

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2002, 21:32:11 ; Search time 2741.47 Seconds  
(without alignments)  
16617.753 Million cell updates/sec

Title: US-09-828-313-22

Sequence: 1 atccgcggcttgatggtc.....gcagatattgccttaacgc 2177

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_ov:\*  
8: gb\_ov:\*  
9: gb\_ov:\*  
10: gb\_ov:\*  
11: gb\_ov:\*  
12: gb\_ov:\*  
13: gb\_ov:\*  
14: gb\_ov:\*  
15: gb\_ov:\*  
16: gb\_ov:\*  
17: gb\_ov:\*  
18: gb\_ov:\*  
19: gb\_ov:\*  
20: gb\_ov:\*  
21: gb\_ov:\*  
22: gb\_ov:\*  
23: gb\_ov:\*  
24: gb\_ov:\*  
25: gb\_ov:\*  
26: gb\_ov:\*  
27: gb\_ov:\*  
28: gb\_ov:\*  
29: gb\_ov:\*  
30: gb\_ov:\*  
31: gb\_ov:\*  
32: gb\_ov:\*  
33: gb\_ov:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	2177	100.0	2177	6	AX280988	AX280988 Sequence
2	540.2	24.8	683	6	AX280975	AX280975 Sequence
3	46.6	2.1	125020	2	AF429315	AF429315 Homo sapi
4	42.8	2.0	157442	2	AC099664	AC099664 Homo sapi
5	42.8	2.0	183978	2	AC024967	AC024967 Homo sapi
6	41.6	1.9	7218	6	166494	166494 Sequence 14
7	40	1.8	70203	2	AC091058	AC091058 Homo sapi
8	40	1.8	153072	2	AL360015	AL360015 Human DNA
9	40	1.8	162508	2	AC025870	AC025870 Homo sapi
10	40	1.8	174200	2	AC022797	AC022797 Homo sapi
11	38.6	1.8	125020	2	AF429315	AF429315 Homo sapi
12	38.6	1.8	155542	30	AC013612	AC013612 Homo sapi
13	38.6	1.8	163325	2	AC010767	AC010767 Homo sapi
14	38.4	1.8	162914	2	AC008813	AC008813 Homo sapi
15	37.8	1.7	69507	2	AC079266	AC079266 Mus muscu
16	37.8	1.7	73465	2	HS256M13	AL109748 Homo sapi
17	37.8	1.7	168182	2	HS214G14	AL050303 Homo sapi
18	37.8	1.7	340000	2	HS21C004	AL163204 Homo sapi
19	37.6	1.7	6706	6	A95311	A95311 Sequence 42
20	37.6	1.7	91940	2	AL445674	AL445674 Human DNA
21	37.6	1.7	157611	2	AC019024	AC019024 Homo sapi
22	37.6	1.7	158307	10	AL592112	AL592112 Mouse DNA
23	37.4	1.7	179726	9	AC007052	AC007052 Homo sapi
24	37.4	1.7	181750	2	AC025506	AC025506 Homo sapi
25	37.4	1.7	192342	2	AC027588	AC027588 Homo sapi
26	37.4	1.7	196869	2	AC087535	AC087535 Homo sapi
27	37.4	1.7	200774	2	AP001592	AP001592 Homo sapi
28	37.2	1.7	155163	2	AC098990	AC098990 Rattus no
29	37.2	1.7	155825	2	AC023503	AC023503 Homo sapi
30	37.2	1.7	164481	2	AC034152	AC034152 Homo sapi
31	37	1.7	173794	9	AC023108	AC023108 Homo sapi
32	37	1.7	176968	9	AL161905	AL161905 Human DNA
33	36.8	1.7	138370	9	AC055707	AC055707 Homo sapi
34	36.8	1.7	152318	2	AL162851	AL162851 Homo sapi
35	36.8	1.7	169141	9	AC021538	AC021538 Homo sapi
36	36.8	1.7	171911	2	AC025637	AC025637 Homo sapi
37	36.6	1.7	62282	9	AC010304	AC010304 Homo sapi
38	36.6	1.7	163167	9	AC011347	AC011347 Homo sapi
39	36.4	1.7	186516	2	AC096938	AC096938 Rattus no
40	36	1.7	119000	9	AP001482	AP001482 Homo sapi
41	36	1.7	126462	9	AC004876	AC004876 Homo sapi
42	36	1.7	167491	2	AC093442	AC093442 Papio cyn
43	36	1.7	168147	9	AL139092	AL139092 Human DNA
44	36	1.7	189058	2	AC013339	AC013339 Homo sapi

#### ALIGNMENTS

RESULT	1	AX280988	Sequence 22 from Patent WO01/7356.	DNA	linear	PAT 02-NOV-2001
LOCUS	AX280988	Sequence 22 from Patent WO01/7356.				
DEFINITION	AX280988	Sequence 22 from Patent WO01/7356.				
ACCESSION	AX280988	GI:16608262				
VERSION	AX280988.1	GI:16608262				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE	1					
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
BASE COUNT	522 a	508 c	547 g	600 t		
ORIGIN						





Db	60619	CTTTGGGAGSCCKKRRKKTSSGVRNRMTSKYTTTKGKTSKSRMGSTMSMGWTCRQYTYTKGTYM	60678
Qy	1713	tgagagcagatctgctcgcgcagcttgcttcataacgcttgcttgctgaaggccctag	1772
Db	60679	WKMKGGSTKRWMYYSKYTCTMTTWTRMMKMKWMMMMTWKYYCMTTKYKKYGGSSSYM	60738
Qy	1773	acggactactcaagaatgaaagatgctgcgtggtgctgctgaagcaactctcttgaggg	1832
Db	60739	YUCNKTMWTCYUWKMKMYRTYKGSMSGCTIMVGRKKMKMMTYMCTTGAMYSIGSAAKGT	60798
Qy	1833	atgctcttcccatctgtagaacaacgcagagcaactctgcgccttgcttgagcaactccscacag	1892
Db	60799	SKAMAKKTTTGCTMDTKMYTKYKNNKMKWBVMKSVYKKRKYVVBDBKNKKKSNHBMKKNND	60858
Qy	1893	aaatatgctcctcaactctggtgtgtcttcscactgcagatacctcaactcctcscactct	1952
Db	60859	DBVAKYBDKTYMBDRNDSDMSRHNHMSMSMDRKYVDMIMRYBBDWYVUNSHNRDNNHNYG	60918
Qy	1953	ctcgcgcctctgcttgagccatcgatcaacacatgcacgcgcacgccttgactgcgctgataa	2012
Db	60919	WKKMKWYIVYHBMYSRWYNDVNDVHMKDKDWCMKSKRSYWKCMGTCMIGKMBNDHDNYM	60978
Qy	2013	tcgtctgtcaatcttagtggagctcttgatgatatagatgcactctgtaagaagcttgga	2072
Db	60979	HDSKGMKCKMKMYGSKSKNNHCMBWMCDBVATWGTRKTCSPBGVSYMTDDNNHNMWB	61038
Qy	2073	tgctgtcttcaagaatgacgcga	2094
Db	61039	WBDSHHMMYDMSMARMAKMGW	61060
RESULT	4		
AC099664/c		157442 bp	DNA
LOCUS			linear HTG 17-NOV-2001
DEFINITION		unm8 sapiens chromosome 3 clone RP11-6B4, WORKING DRAFT SEQUENCE, 2	
ACCESSION		AC099664.1	GI:16973745
VERSION		AC099664.1	GI:16973745
KEYWORDS		HTG; HTGS-PHASE1; HTGS-DRAFT; HTGS-FULLTOP; HTGS-ACTIVEFIN.	
KEIMOROS		human.	
SOURCE		Human sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE		Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.	
AUTHORS		Direct Submission	
REFERENCE		Submitted (17-NOV-2001) Genome Center, University of Washington, Box 353115, Seattle, WA 98195, USA	
AUTHORS		On Nov 17, 2001 this sequence version replaced gi:13194374.	
JOURNAL			
TITLE			
COMMENT			
		Genome Center	
		Center: University of Washington Genome Center	
		Center Code: UWGC	
		Web site: http://www.genome.washington.edu	
		Contact: uwgchgs@u.washington.edu	
		Drafting Center: BCM	
		Project Information	
		Center project name: chr-3	
		Center clone name: RP11-6B4 (bc0602)	
		Summary Statistics	
		Sequencing vector: unknown; 44% of reads	
		Sequencing vector: plasmid; L08752; 56% of reads	
		Chemistry: dye-terminator EM; 89% of reads	
		Chemistry: dye-terminator Big Dye; 11% of reads	
		Assembly program: Phrap; version 0.990319	
		Consensus quality: 156673 bases at least 040	
		Consensus quality: 156673 bases at least 030	
		Consensus quality: 157102 bases at least 020	
		Insert size: 176750; 14.0% error; agarose-1p	
		Insert size: 157342; sum-of-confliqs	

```

Quality coverage: 6.9x in Q20 bases; agarose-gel
Quality coverage: 7.8x in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 65404: contig of 65404 bp in length
* 65405 65504: gap of unknown length
* 65505 157442: contig of 91938 bp in length.
FEATURES
Source
1..157442 Homo sapiens
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-6B4"
/clone_11b="RPC1 human BAC library 11"
1..65404
misc-feature
/note="assembly_name:Contig35"
65505..157442
/misc-feature
/note="assembly_name:Contig36"
BASE COUNT 50491 a 28472 c 27460 g 50919 t 100 others
ORIGIN
Query Match 2.0%; Score 42.8; DB 2; Length 157442;
Best Local Similarity 48.4%; Pred. No. 0.48;
Matches 119; Conservative 0; Mismatches 127; Indels 0; Gaps 0
OY 1891 agaaacatagcttaccatcttggtggtgtccacccatggatcatatgacatcttcata 1950
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49206 AGAAGTACTACTGATGATTTGTTATTTCTCTCGTTATTTGTTATGTCCTTTATTCCTC 49147
OY 1951 ttctcgagcttgcttcagatcatalacacatgcagctacgtcacttgcagctgagat 2010
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49146 TCTTGGTGTCTCTCTTGGAGATTAAATTTTGGCTAGTAGAGATTCTATGATTTTACTTTT 49087
OY 2011 aatcgcttgcatttgatgagacttgatgataagatgcacattgtacagtagcttg 2070
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49086 TATCTTTGTGTGATATCTATAGCTTTTGGCTTTGTGTTACCATGAGCTTACATAAA 49027
OY 2071 gatgctgtttacaagaatagcgcagctagaaagccttaaccctttagctaccatgatatt 2130
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49026 CATGCTATCTATTAATAAGGGGCGCTTTACACCAATATCAACTTACTTGATTTGCATATA 48967
OY 2131 ttaaac 2136
|||||
Db 48966 ATTAATC 48961
RESULT 5
AC024967
LOCUS Homo sapiens chromosome 3 clone RP11-554B18 map 3, WORKING DRAFT
DEFINITION AC024967
AC024967
AC024967.4 GI:13959248
VERSION AC024967.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 183978)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
JOURNAL Homo sapiens chromosome 3, clone RP11-554B18
REFERENCE 2 (bases 1 to 183978)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,T., Barna,N., Bastien,V., Beda,F.,

```





```

* 6817 7597: contig of 781 bp in length
* 7598 7697: gap of 100 bp
* 8450 8449: contig of 752 bp in length
* 8550 9312: contig of 763 bp in length
* 9313 9412: gap of 100 bp
* 10175 10174: contig of 762 bp in length
* 10275 11033: contig of 759 bp in length
* 11034 11133: gap of 100 bp
* 11134 11910: contig of 777 bp in length
* 12011 12722: contig of 712 bp in length
* 12723 12832: gap of 100 bp
* 12833 13529: contig of 707 bp in length
* 13530 13629: gap of 100 bp
* 13630 14358: contig of 729 bp in length
* 14359 14458: gap of 100 bp
* 14459 15238: contig of 780 bp in length
* 15239 15338: gap of 100 bp
* 15339 16049: contig of 711 bp in length
* 16050 16149: gap of 100 bp
* 16150 16903: contig of 754 bp in length
* 16904 17004: gap of 100 bp
* 17004 17552: contig of 749 bp in length
* 17553 17852: gap of 100 bp
* 17853 18576: contig of 724 bp in length
* 18577 18676: gap of 100 bp
* 18677 19395: contig of 719 bp in length
* 19396 19495: gap of 100 bp
* 19496 20253: contig of 758 bp in length
* 20254 20353: gap of 100 bp
* 20354 21099: contig of 746 bp in length
* 21100 21199: gap of 100 bp
* 21200 21917: contig of 718 bp in length
* 21918 22017: gap of 100 bp
* 22018 22687: contig of 670 bp in length
* 22688 22787: gap of 100 bp
* 22788 23498: contig of 711 bp in length
* 23499 23598: gap of 100 bp
* 23599 24318: contig of 720 bp in length
* 24319 24418: gap of 100 bp
* 24419 25180: contig of 762 bp in length
* 25181 25280: gap of 100 bp
* 25281 26009: contig of 729 bp in length
* 26010 26109: gap of 100 bp
* 26110 26865: contig of 756 bp in length
* 26866 26965: gap of 100 bp
* 26966 27274: contig of 780 bp in length
* 27275 27845: gap of 100 bp
* 27846 28556: contig of 711 bp in length
* 28557 28656: gap of 100 bp
* 28657 29414: contig of 758 bp in length
* 29415 29514: gap of 100 bp
* 29515 30240: contig of 726 bp in length
* 30241 30340: gap of 100 bp
* 30341 31109: contig of 769 bp in length
* 31110 31209: gap of 100 bp
* 31210 31928: contig of 720 bp in length
* 31930 32028: gap of 100 bp
* 32030 32752: contig of 723 bp in length
* 32753 32852: gap of 100 bp
* 32853 33553: contig of 701 bp in length
* 33554 33653: gap of 100 bp
* 33654 34362: contig of 709 bp in length
* 34363 34462: gap of 100 bp
* 34463 35226: contig of 764 bp in length
* 35227 35326: gap of 100 bp
* 35327 36088: contig of 752 bp in length
* 36089 36188: gap of 100 bp
* 36189 36927: contig of 739 bp in length
* 36928 37027: gap of 100 bp
* 37028 37761: contig of 734 bp in length

```

```

* 37762 37861: gap of 100 bp
* 37862 38616: contig of 755 bp in length
* 38617 38716: gap of 100 bp
* 38717 39475: contig of 759 bp in length
* 39476 39575: gap of 100 bp
* 39576 40329: contig of 754 bp in length
* 40330 40429: gap of 100 bp
* 40430 41154: contig of 725 bp in length
* 41155 41254: gap of 100 bp
* 41255 41987: contig of 733 bp in length
* 41988 42087: gap of 100 bp
* 42088 42805: contig of 718 bp in length
* 42806 42905: gap of 100 bp
* 42906 43656: contig of 751 bp in length
* 43657 43756: gap of 100 bp
* 43757 44475: contig of 719 bp in length
* 44476 44575: gap of 100 bp
* 44576 45313: contig of 738 bp in length
* 45314 45413: gap of 100 bp
* 45414 46145: contig of 732 bp in length
* 46146 46245: gap of 100 bp
* 46246 46996: contig of 751 bp in length
* 46997 47096: gap of 100 bp
* 47097 47833: contig of 737 bp in length
* 47834 47933: gap of 100 bp
* 47934 48655: contig of 722 bp in length
* 48656 48755: gap of 100 bp
* 48756 49474: contig of 719 bp in length
* 49475 49574: gap of 100 bp
* 49575 50310: contig of 736 bp in length
* 50311 50410: gap of 100 bp
* 50411 51135: contig of 725 bp in length
* 51136 51235: gap of 100 bp
* 51236 51944: contig of 709 bp in length
* 51945 52044: gap of 100 bp
* 52045 52755: contig of 711 bp in length
* 52756 52855: gap of 100 bp
* 52856 53627: contig of 772 bp in length
* 53628 53727: gap of 100 bp
* 53728 54482: contig of 755 bp in length
* 54483 54582: gap of 100 bp
* 54583 55325: contig of 743 bp in length
* 55326 55425: gap of 100 bp
* 55426 56168: contig of 743 bp in length
* 56169 56268: gap of 100 bp
* 56269 56981: contig of 713 bp in length
* 56982 57081: gap of 100 bp
* 57082 57844: contig of 763 bp in length
* 57845 57944: gap of 100 bp

```

Query Match 1.8%; Score 40; DB 2; Length 70203;  
 Best Local Similarity 58.3%; Pred. No. 3.1;  
 Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```

Oy 1897 tagttcttaacttcagggttcgttcaccactagaccatagaccatcatatttcg 1956
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30240 TATCTCTTCAATTGACCTTTCCACAAATATGAATGACCTCTCTTTTAA 30181

Oy 1957 gacattgttcgagctacatgaccatcagcgcacattgacgcgtgatatcgt 2016
||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30180 ATTAAGTTTATTAGCTCAATTATCTGTGACAAAGATCACTTCGACGCTTGGAATTGT 30121

```

RESULT 8  
 AL360015/c  
 LOCUS Human DNA sequence from clone Rpl1-344P20 on chromosome 6, complete  
 DEFINITION  
 accession AL360015.25 GI:12539717  
 VERSION  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 153072)

Wall, M.

Direct Submission

Submitted (22-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

request: clonerequests@sanger.ac.uk

On Jan 26, 2001 this sequence version replaced g1:12227402.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated repeat sequence elements. Where the sequence is  
ambiguous, there is an annotation using the 'unsure' feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr6>

Rp11-344F20 is from the library RCI-11.2 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone  
Rp11-344F20 it may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.

The true right end of clone Rp11-344F20 is at 153072 in this  
sequence. The true right end of clone Rp11-505K1 is at 100 in this  
sequence.

## FEATURES

## source

1. 153072  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="Rp11-344F20"  
/clone\_11b="RCI-11.2"  
60. 177  
/note="MIR repeat: matches 28. .144 of consensus"  
714. 875  
/note="MIR repeat: matches 80. .252 of consensus"  
1068. 1143  
/note="L2 repeat: matches 2641. .2717 of consensus"  
3293. 3530  
/note="MER8 repeat: matches 1. .239 of consensus"  
4107. 4180  
/note="37 copies 2 mer ac 91% conserved"  
4421. 4732  
/note="Alu repeat: matches 1. .310 of consensus"  
4736. 5207  
/note="MLTII repeat: matches 33. .541 of consensus"  
6051. 6176  
/note="63 copies 2 mer aa 59% conserved"  
6629. 7114  
/note="L2 repeat: matches 2238. .2746 of consensus"  
7120. 7215  
/note="48 copies 2 mer ta 69% conserved"  
8053. 8404  
/note="MLTII repeat: matches 12. .365 of consensus"  
8872. 9833  
/note="LIPAS repeat: matches 5184. .6143 of consensus"  
11814. 14211  
/note="LIPAT repeat: matches 3692. .6120 of consensus"  
15010. 15150

repeat\_region  
/note="MIR repeat: matches 73. .218 of consensus"  
15163. 15543  
/note="MIR repeat: matches 1. .426 of consensus"  
15800. 15900  
/note="MER33 repeat: matches 33. .135 of consensus"  
15903. 16040  
/note="L2 repeat: matches 2183. .2324 of consensus"  
16259. 16489  
/note="L2 repeat: matches 1818. .2033 of consensus"  
16560. 16623  
/note="MER96 repeat: matches 1. .60 of consensus"  
16918. 17030  
/note="MER91C repeat: matches 1. .119 of consensus"  
17234. 17285  
/note="26 copies 2 mer ag 88% conserved"  
17305. 17526  
/note="MIR repeat: matches 22. .258 of consensus"  
18236. 18606  
/note="LIMD2 repeat: matches 5494. .5968 of consensus"  
18607. 18978  
/note="LIMD2 repeat: matches 5951. .6321 of consensus"  
20104. 20471  
/note="LIPAT2 repeat: matches 4063. .4405 of consensus"  
20472. 20828  
/note="THE1C repeat: matches 3. .371 of consensus"  
20830. 22459  
/note="THE1C-internal repeat: matches 4. .1578 of  
consensus"  
22463. 22827  
/note="THE1C repeat: matches 1. .371 of consensus"  
22828. 24496  
/note="LIPAT2 repeat: matches 4405. .6153 of consensus"  
24508. 24802  
/note="Alu repeat: matches 1. .285 of consensus"  
25079. 25177  
/note="U6 repeat: matches 1. .99 of consensus"  
25187. 25625  
/note="LIP repeat: matches 2128. .2598 of consensus"  
25614. 25970  
/note="LIPAT5 repeat: matches 5128. .5472 of consensus"  
25971. 26271  
/note="Alu repeat: matches 1. .300 of consensus"  
26272. 26917  
/note="LIPAT5 repeat: matches 5472. .6153 of consensus"  
27112. 27147  
/note="18 copies 2 mer tt 88% conserved"  
27433. 27482  
/note="25 copies 2 mer tt 76% conserved"  
27489. 28511  
/note="LIM8 repeat: matches 5254. .6288 of consensus"  
28518. 28859  
/note="THE1B repeat: matches 3. .364 of consensus"  
28860. 30410  
/note="THE1B-internal repeat: matches 1. .1580 of  
consensus"  
30416. 30731  
/note="THE1B repeat: matches 15. .364 of consensus"  
30733. 31170  
/note="LIM8 repeat: matches 4823. .5253 of consensus"  
31622. 31890  
/note="LIMC repeat: matches 1738. .1988 of consensus"  
31891. 32097  
/note="MER6B repeat: matches 2. .208 of consensus"  
32098. 32360  
/note="LIMC repeat: matches 1988. .2256 of consensus"  
33222. 33283  
/note="MER96 repeat: matches 115. .172 of consensus"  
34160. 34490  
/note="MER7A repeat: matches 1. .346 of consensus"  
38388. 38597  
/note="MIR repeat: matches 31. .250 of consensus"  
41237. 41312  
/note="MIR repeat: matches 49. .124 of consensus"





```
* 21164 21263: gap of 100 bp
* 21264 22654: contig of 1391 bp in length
* 22655 22754: gap of 100 bp
* 22755 24290: contig of 1536 bp in length
* 24291 24390: gap of 100 bp
* 24391 25594: contig of 1204 bp in length
* 25595 25694: gap of 100 bp
* 25695 27397: contig of 1703 bp in length
* 27398 27497: gap of 100 bp
* 27498 29330: contig of 1833 bp in length
* 29331 29430: gap of 100 bp
* 29431 31357: contig of 1927 bp in length
* 31358 31457: gap of 100 bp
* 31458 33360: contig of 1903 bp in length
* 33361 33460: gap of 100 bp
* 33461 36064: contig of 2604 bp in length
* 36065 36164: gap of 100 bp
* 36165 37979: contig of 1815 bp in length
* 37980 38079: gap of 100 bp
* 38080 40463: contig of 2384 bp in length
* 40464 40563: gap of 100 bp
* 40564 43574: contig of 3011 bp in length
* 43575 43674: gap of 100 bp
* 43675 45259: contig of 1585 bp in length
* 45260 45359: gap of 100 bp
* 45360 47385: contig of 2026 bp in length
* 47386 47485: gap of 100 bp
* 47486 49679: contig of 2194 bp in length
* 49680 49779: gap of 100 bp
* 49780 51819: contig of 2040 bp in length
* 51820 51919: gap of 100 bp
* 51920 55270: contig of 3351 bp in length
* 55271 55370: gap of 100 bp
* 55371 58223: contig of 2853 bp in length
* 58224 58323: gap of 100 bp
* 58324 61669: contig of 3246 bp in length
* 61670 61669: gap of 100 bp
* 61670 65776: contig of 4107 bp in length
* 65777 65876: gap of 100 bp
* 65877 68164: contig of 2288 bp in length
* 68165 68264: gap of 100 bp
* 68265 72330: contig of 4066 bp in length
* 72331 72430: gap of 100 bp
* 72431 75721: contig of 3291 bp in length
* 75722 75821: gap of 100 bp
* 75822 79327: contig of 3506 bp in length
* 79328 79427: gap of 100 bp
* 79428 86170: contig of 6743 bp in length
* 86171 86270: gap of 100 bp
* 86271 92478: contig of 6208 bp in length
* 92479 92578: gap of 100 bp
* 92579 98667: contig of 6089 bp in length
* 98668 98767: gap of 100 bp
* 98768 105029: contig of 6262 bp in length
* 105030 105129: gap of 100 bp
* 105130 111321: contig of 6192 bp in length
* 111322 111421: gap of 100 bp
* 111422 117477: contig of 6056 bp in length
* 117478 117577: gap of 100 bp
* 117578 125755: contig of 8178 bp in length
* 125756 125855: gap of 100 bp
* 125856 133343: contig of 7488 bp in length
* 133344 133443: gap of 100 bp
* 133444 141679: contig of 8236 bp in length
* 141680 141779: gap of 100 bp
* 141780 153666: contig of 11887 bp in length
* 153667 153766: gap of 100 bp
* 153767 162034: contig of 8268 bp in length
* 162035 162134: gap of 100 bp
* 162135 162508: contig of 374 bp in length.
Location/Qualifiers
1..162508
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone="RP11-25P4"
/clone_id="RPC1-11 Human Male BAC"
1..1133
misc_feature
/note="assembly_fragment"
1234..2439
misc_feature
/note="assembly_fragment"
2540..4058
misc_feature
/note="assembly_fragment"
4159..5400
misc_feature
/note="assembly_fragment"
5501..18786
misc_feature
/note="assembly_fragment"
18887..19943
misc_feature
/note="assembly_fragment"
20044..21163
misc_feature
/note="assembly_fragment"
21264..22654
misc_feature
/note="assembly_fragment"
22755..24290
misc_feature
/note="assembly_fragment"
24391..25594
misc_feature
/note="assembly_fragment"
25695..27397
misc_feature
/note="assembly_fragment"
27498..29330
misc_feature
/note="assembly_fragment"
29431..31357
misc_feature
/note="assembly_fragment"
31458..33360
misc_feature
/note="assembly_fragment"
33461..36064
misc_feature
/note="assembly_fragment"
36165..37979
misc_feature
/note="assembly_fragment"
43675..45259
misc_feature
/note="assembly_fragment"
45360..47385
misc_feature
/note="assembly_fragment"
Query Match 1.8%, Score 40; DB 2: Length 162508;
Best Local Similarity 58.3%, Pred. No. 3.6;
Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 1897 tagagcttcacatcgtggtctgtcaccacagatcatatgacatcctcattatctcg 1956
||| ||||| || | || | ||| ||||| || ||| |||
DB 88829 TACTCCTTCAMTTTGAGCTTCTTCACAAATGAATGACCTCTCTTTTAA 88888
QY 1957 ggccttcttcgagcctatgacatcatgacatccttgcacttgatcgagtgatcatct 2016
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 88889 ATTATGCTTTTACCTCAATATCTGTGACAGATGACTTGCGACGCTTGAGGAATGT 88948
RESULT 10
AC022797 174200 bp DNA linear HTG 16-MAR-2000
LOCUS Homo sapiens chromosome 6 clone RP11-281G13 map 6, WORKING DRAFT
DEFINITION SHOUNCE, 22 unordered pieces.
ACCESSION AC022797.3 GI:7249244
VERSION AC022797
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 174200)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 6, clone RP11-281G13
JOURNAL Unpublished
```

REFERENCE  
AUTHORS

2 (bases 1 to 174200)

Allen, N.  
Anderson, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,  
Britten, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
Bouslavsky, I., Boukgalter, B., Brown, A., Burkett, G., Castle, A.,  
Chopel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fennestor, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardya, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howard, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Landers, T., Lehotsky, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
Macdonald, P., Margulis, N., Mckean, P., McGarr, A., McKernan, K.,  
McPheeters, R., Meldrum, J., Menelus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
Pietre, N., Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Tamas, J., Testate, S., Theodore, J.,  
Tirrell, A., Vassiliou, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W., J.,  
Zimmer, A. and Zody, W.

TITLE  
JOURNAL

Direct Submission

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2000 this sequence version replaced gi:5984458.

All repeats were identified using RepeatMasker:

SMIT, A.F.A. & Green, P. (1996-1997)

<http://ltp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

## ----- Project Information -----

Center project name: L6043

Center clone name: 281\_G\_13

----- Summary Statistics -----

Sequencing vector: M13: M77815: 100% of reads

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phran: version 0 950731

Consensus quality: 163082 bases at 100% MQ

consensus quality: 100002 bases at 10000 Q40  
consensus quality: 168787 bases at 10000 Q30

Consensus quality: 10678 bases at least Q30  
Consensus quality: 170707 bases at least Q30

consensus quality: 1/0/9/ bases

Insert size: 182000; agarose-1p

Insert size: 172100; sum-of-configs

Quality coverage: 3.6 in Q20 bases; agarose-tp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 config. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the configs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

*	1	3194	: contig of 3194 bp	In length
*	3195	3294	: gap of 100 bp	
*	3295	5794	: contig of 2500 bp	In length
*	5795	5894	: gap of 100 bp	
*	5895	9861	: contig of 3967 bp	In length
*	9862	9961	: gap of 100 bp	
*	9962	13464	: contig of 3503 bp	In length
*	13465	13564	: gap of 100 bp	
*	13565	15205	: contig of 1641 bp	In length
*	15206	15305	: gap of 100 bp	
*	15306	19909	: contig of 4604 bp	In length
*	19910	20009	: gap of 100 bp	
*	20010	23963	: contig of 3954 bp	In length
*	23964	24063	: gap of 100 bp	
*	24064	29018	: contig of 4955 bp	In length
*	29019	29118	: gap of 100 bp	
*	29119	34619	: contig of 5501 bp	In length
*	34620	34719	: gap of 100 bp	
*	34720	42600	: contig of 7341 bp	In length
*	42601	42160	: gap of 100 bp	

## FEATURES

## SOURCE

*	42161	48254:	contig of 6094 bp	in length
*	48255	48354:	gap of 100 bp	
*	48355	55799:	contig of 7745 bp	in length
*	55800	55899:	gap of 100 bp	
*	55900	64293:	contig of 8394 bp	in length
*	64294	64393:	gap of 100 bp	
*	64394	70209:	contig of 5816 bp	in length
*	70210	70309:	gap of 100 bp	
*	70310	80175:	contig of 9866 bp	in length
*	80176	80275:	gap of 100 bp	
*	80276	90721:	contig of 10446 bp	in length
*	90722	90821:	gap of 100 bp	
*	90822	103432:	contig of 12611 bp	in length
*	103433	103532:	gap of 100 bp	
*	103533	116657:	contig of 13125 bp	in length
*	116658	116757:	gap of 100 bp	
*	116758	128699:	contig of 11942 bp	in length
*	128700	128799:	gap of 100 bp	
*	128800	143088:	contig of 14289 bp	in length
*	143089	143188:	gap of 100 bp	
*	143189	157620:	contig of 14432 bp	in length
*	157621	157720:	gap of 100 bp	
*	157721	174200:	contig of 16450 bp	in length

Location/Qualifiers

source	1. 11/4/2001
misc_feature	/organism="Homo sapiens"
misc_feature	/db_xref="taxon:9606"
misc_feature	/chromosome="6"
misc_feature	/map="6"
misc_feature	/clone="RP11-281G13"
misc_feature	/clone_id="RPCL11 Human
misc_feature	1. 3194
misc_feature	/note="assembly-fragment"
misc_feature	3295. 5794
misc_feature	/note="assembly-fragment"
misc_feature	5895. 9861
misc_feature	/note="assembly-fragment"
misc_feature	9962. 13464
misc_feature	/note="assembly-fragment"
misc_feature	13565. 15205
misc_feature	/note="assembly-fragment"
misc_feature	clone_end:5P6
misc_feature	vector_side:left"
misc_feature	15306. 19909
misc_feature	/note="assembly-fragment"
misc_feature	20010. 23963
misc_feature	/note="assembly-fragment"
misc_feature	24064. 29018
misc_feature	/note="assembly-fragment"
misc_feature	2919. 34619
misc_feature	/note="assembly-fragment"
misc_feature	34720. 42060
misc_feature	/note="assembly-fragment"
misc_feature	42161. 48254
misc_feature	/note="assembly-fragment"
misc_feature	48355. 55799
misc_feature	/note="assembly-fragment"
misc_feature	55900. 64293
misc_feature	/note="assembly-fragment"
misc_feature	64394. 70209
misc_feature	/note="assembly-fragment"
misc_feature	70310. 80175
misc_feature	/note="assembly-fragment"
misc_feature	80276. 90721
misc_feature	/note="assembly-fragment"
misc_feature	90822. 103432
misc_feature	/note="assembly-fragment"
misc_feature	103533. 116657
misc_feature	/note="assembly-fragment"
misc_feature	clone_end:17
misc_feature	vector_side:left"
misc_feature	116758. 128659
misc_feature	/note="assembly-fragment"



Db 50961 ARGBGMKCYRTSCMKRYYYMHKSSHDSBGYKSKRSYBYWTCOMKMSVDSHMRD 50902  
 QY 1104 ttgcgaagagcgctagcgagcgaagtgttcgttggaagccggtttgcagagcatctccg 1163  
 Db 50901 MMTDKMRKVSSTGRCYHBSHSMCKSMRSRSDSGYGMTHMYSKSSASAHMYISGMK 50842  
 QY 1164 agatgaagcctcgtcatcgaattccacactccgacgtcgtacggagatggtgtctccgc 1223  
 Db 50841 SSVYBMDMSMYTBSKSYMBRMSGSYBRYRCAYMRRYBHRSSRSTRSMYGGKRCW 50782  
 QY 1224 cgaagaacgcctc 1236  
 Db 50781 RDSYBRRSRVHTS 50769

RESULT 12  
 AC013612/c standard; DNA; HTG; 155542 BP.  
 ID AC013612  
 XX AC013612;  
 XX AC013612.3  
 XX  
 SV 15-NOV-1999 (Rel. 61, Created)  
 DT 03-APR-2000 (Rel. 63, Last updated, Version 3)  
 XX  
 DE Homo sapiens clone RP11-10L7, WORKING DRAFT SEQUENCE, 10 unordered pieces.  
 XX  
 KW HTG: HTGS\_DRAFT; HTGS\_PHASE1.  
 OS Homo sapiens (human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 XX  
 RN [1]  
 RP 1-155542  
 RA Birren B., Linton L., Nusbaum C., Lander E.:  
 RT Homo sapiens, clone RP11-10L7;  
 RL unpublished.  
 XX  
 XX  
 RN [2]  
 RP 1-155542  
 RA Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,  
 RA Baldwin J., Barna N., Beckerly R., Boguslavsky L., Boukhgalter B.,  
 RA Brown A., Castle A., Colangelo M., Collins S., Collymore A., Cooke P.,  
 RA Deceallano K., Dewar K., Domingo M., Donelan L., Doyle M., Ferreira P.,  
 RA Fitzhugh W., Forrest C., Funke R., Gage D., Galagan J., Gardyna S.,  
 RA Grant G., Hagos B., Heatford A., Horton L., Howland J.C., Johnson R.,  
 RA Jones C., Kann L., Karatas A., Klein J., Lehoczy J., Lile C., Locke K.,  
 RA MacDonald P., Marquis N., McEwan P., McGuire C.H., McKernan K., McLaughlin J.,  
 RA Meldrum J., Morrow J., Naylor J., Norman C.H., O'Connor T., O'Donnell P.,  
 RA Peterson K., Pollara V., Riley R., Roy A., Santos R., Severy P.,  
 RA Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J., Testaye S.,  
 RA Tirrell A., Vassiliev H., Vo A., Wheeler J., Wu X., Wyman D., Ye W.J.,  
 RA Zimmer A., Zody M.;  
 RT  
 RL Submitted (13-NOV-1999) to the EMBL/Genbank/DBJ databases.  
 RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,  
 RL Cambridge, MA 02141, USA  
 RL  
 XX  
 XX  
 CC On Mar 30, 2000 this sequence version replaced gi:6532098.  
 CC All repeats were identified using RepeatMasker:  
 CC Shult, A.F.A. & Green, P. (1996-1997)  
 CC http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 CC ----- Genome Center  
 CC Center: Whitehead Institute/ MIT Center for Genome Research  
 CC Center code: WIBR  
 CC Web site: http://www-seq.wi.mit.edu  
 CC Contact: sequence\_submissions@genome.wi.mit.edu  
 CC ----- Project Information  
 CC Center project name: L3101  
 CC Center clone name: 10\_L\_7  
 CC ----- Summary Statistics

CC Sequencing vector: M13; M77815; 100% of reads  
 CC Chemistry: Dye-terminator Big Dye; 100% of reads  
 CC Assembly program: Phrap; version 0.960731  
 CC Consensus quality: 151809 bases at least Q40  
 CC Consensus quality: 153486 bases at least Q30  
 CC Consensus quality: 154041 bases at least Q20  
 CC Insert size: 157000; agarose-fp  
 CC Insert size: 154642; sum-of-contigs  
 CC Quality coverage: 5.5 in Q20 bases; agarose-fp  
 CC Quality coverage: 5.6 in Q20 bases; sum-of-contigs  
 CC -----  
 CC \* NOTE: This is a 'working draft' sequence. It currently  
 CC \* consists of 10 contigs. The true order of the pieces  
 CC \* is not known and their order in this sequence record is  
 CC \* arbitrary. Gaps between the contigs are represented as  
 CC \* runs of N, but the exact sizes of the gaps are unknown.  
 CC \* This record will be updated with the finished sequence  
 CC \* as soon as it is available and the accession number will  
 CC \* be preserved.  
 CC \*  
 CC 1 3918: contig of 3918 bp in length  
 CC \* 3919 4018: gap of 100 bp  
 CC \* 4019 7852: contig of 3834 bp in length  
 CC \* 7853 7952: gap of 100 bp  
 CC \* 7953 10420: contig of 2468 bp in length  
 CC \* 10421 10520: gap of 100 bp  
 CC \* 10521 16131: contig of 5611 bp in length  
 CC \* 16132 16231: gap of 100 bp  
 CC \* 16232 21713: contig of 5482 bp in length  
 CC \* 21714 21813: gap of 100 bp  
 CC \* 21814 25983: contig of 4170 bp in length  
 CC \* 25984 26083: gap of 100 bp  
 CC \* 26084 33604: contig of 7521 bp in length  
 CC \* 33605 33704: gap of 100 bp  
 CC \* 33705 45042: contig of 11338 bp in length  
 CC \* 45043 45142: gap of 100 bp  
 CC \* 45143 77339: contig of 32197 bp in length  
 CC \* 77340 77439: gap of 100 bp  
 CC \* 77440 135542: contig of 78103 bp in length.  
 CC  
 XX  
 FH Key Location/Qualifiers  
 FH  
 FT source  
 FT 1. 155542  
 FT /db\_xref="taxon:9606"  
 FT /organism="Homo sapiens"  
 FT /clone="RP11-10L7"  
 FT /clone\_1b="RP11-11 Human Male BAC"  
 FT  
 FT misc\_feature  
 FT 1. 3918  
 FT /note="assembly-fragment"  
 FT 4019..7852  
 FT /note="assembly-fragment"  
 FT 7953..10420  
 FT /note="assembly-fragment"  
 FT 10521..16131  
 FT /note="assembly-fragment"  
 FT 16232..21713  
 FT /note="assembly-fragment"  
 FT 21814..25983  
 FT /note="assembly-fragment-clone\_end:r7-vector\_side:left"  
 FT 26084..33604  
 FT /note="assembly-fragment"  
 FT 33705..45042  
 FT /note="assembly-fragment"  
 FT 45143..77339  
 FT /note="assembly-fragment-clone\_end:SP6-vector\_side:right"  
 FT 77440..155542  
 FT /note="assembly-fragment"  
 FT  
 XX  
 SO Sequence 155542 BP; 44570 A; 32359 C; 30817 G; 46895 T; 901 other;

Query Match 1.88; Score 38.6; DB 30; Length 155542;  
 Best Local Similarity 51.4%; Pred. No. 9.7;  
 Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
OY 305 gggtcagagatcatgatacgcctcagcgagagaagagccctgactgttcccgagcgtac 364
||||| 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 108821 GGGTGCAGCGGGGCTGAGTCCGAAAGAAATCAGCAAGAAAGATAGGGGTGGGCGACGTT 108762

OY 365 gctgaggattgactgaagaagaggtgagccgacatcaactattatctctgtgaggga 424
||||| 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 108761 TTATNAGATTGGGTAGCTACGACGAAATTTACAGTCAAGGAGGGGTTGTTCTTGAGGGG 108702

OY 425 agagtcaggttccaatgggtccacccgttcttctgtgtcaggttcaagg 477
||||| 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 108701 AGGCTTAGGCTCAGCAAGTGTCTAGTGGGAGCTTCTTGACCGACGAGAAG 108649

RESULT 13
AC107067/c 156325 bp DNA linear HTG 07-FEB-2002
LOCUS
DEFINITION Homo sapiens chromosome 4 clone RP11-10L7, WORKING DRAFT SEQUENCE,
ACCESSION AC107067
VERSION AC107067.2 GI:1854834
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Materston R.H.
1 (bases 1 to 156325)
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 156325)
Materston, R.H.
Direct Submission
Submitted (15-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 7, 2002 this sequence version replaced gi:18151028.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Drafting Center: WIBR

----- Summary Statistics -----
Sequencing vector: M13, 53%
Sequencing method: plasmid, 47%
Chemistry: Dye-terminator ET, 0% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 156775 bases at least Q40
Consensus quality: 157110 bases at least Q30
Consensus quality: 157257 bases at least Q20
Insert size: 153000; agarose-fp
Insert size: 157371; sum-of-coverage
Quality coverage: 10.23 in Q20 bases; sum-of-coverage
Quality coverage: 10.21 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 156325: contig of 156325 bp in length.
Location/Qualifiers
1..156325
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-10L7"
misc_feature 1..156325
BASE COUNT 45699 a 32863 c 31085 g 46678 t
ORIGIN

Query Match 1.88; Score 38.6; DB 2; Length 156325;
Best Local Similarity 51.4%; Pred. No. 9.7;
Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 305 gggtcagagatcatgatacgcctcagcgagagaagagccctgactgttcccgagcgtac 364
||||| 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 70586 GGGTGCAGCGGGGCTGAGTCCGAAAGAAATCAGCAAGAAAGATAGGGGTGGGCGACGTT 70527

OY 365 gctgaggattgactgaagaagaggtgagccgacatcaactattatctctgtgaggga 424
||||| 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 70526 TTATNAGATTGGGTAGCTACGACGAAATTTACAGTCAAGGAGGGGTTGTTCTTGAGGGGT 70467

OY 425 agagtcaggttccaatgggtccacccgttcttctgtgtcaggttcaagg 477
||||| 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 70466 AGGCTTAGGCTCAGCAAGTGTCTAGTGGGAGCTTCTTGACCGACGAGAAG 70414

RESULT 14
AC008813 162914 bp DNA linear PRI 26-NOV-2000
LOCUS
DEFINITION Homo sapiens chromosome 19 clone CTD-2102P23, complete sequence.
ACCESSION AC008813
VERSION AC008813.6 GI:11344986
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Materston R.H.
1 (bases 1 to 162914)
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 162914)
Materston, R.H.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 162914)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (26-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Nov 26, 2000 this sequence version replaced gi:8575922.
Draft sequence produced by DOE Joint Genome Institute

www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.9.
STS Content:
SHGC-56762 G37043
SHGC-57254 G38378
WT-6292 G06202
SHGC-89248 G61707
WT-11361 G24462.

Location/Qualifiers
1..162914
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-2102P23"
BASE COUNT 42170 a 36675 c 37407 g 46662 t
ORIGIN
```



Mon Sep 16 08:49:18 2002

us-09-828-313-22.rge

Page 16

[illegible]

Search completed: September 15, 2002, 23:57:46  
Job time: 8735 sec

•  
•  
•  
•





PI Da Costa Silva EO, Bohmert HJ, Van Thiejen N, Chen R;  
PI Sarria-Mallén R;  
XX WPI: 2002-049153/06.  
DR P-PSDB: AAM52838.  
XX  
PT New protein, useful for increasing tolerance to environmental stress,  
PT comprises a protein kinase stress-related protein selected from  
PT protein kinases, casein kinase homologs, MAP kinases or calcium  
PT dependent protein kinases  
PS Claim 14: Fig 21: 154bp: English.  
XX  
CC Sequences AAM52830-AAM52842 represent novel protein kinase stress-related  
CC proteins (PKSRPs) from the moss *Physcomitrella patens*, and sequences  
CC ABA91069-ABA91081 represent full-length cDNAs encoding them. The cDNA  
CC sequences were obtained from expressed sequence tags (ESTs: ABA91056-  
CC ABA91068) derived from *Physcomitrella patens* cDNA libraries. The PKSRPs  
CC of the invention comprise protein kinase-6 (PK-6), protein kinase-7  
CC (PK-7), protein kinase-8 (PK-8), protein kinase-9 (PK-9), casein kinase  
CC homologue-1 (CK-1), casein kinase homologue-2 (CK-2), casein kinase  
CC homologue-3 (CK-3), mitogen-activated protein (MAP) kinase-2 (MK-2),  
CC MAP kinase-3 (MK-3), MAP kinase-4 (MK-4), MAP kinase-5 (MK-5),  
CC calcium-dependent protein kinase-1 (CPK-1), and calcium-dependent protein  
CC kinase-2 (CK-2). When overexpressed, the PKSRPs are able to confer  
CC tolerance to environmental stresses such as salinity, drought,  
CC temperature, metal, chemical, pathogenic and oxidative stress.  
CC *Physcomitrella patens* PKSRP nucleic acids may be used to generate  
CC transgenic plants and seeds with increased tolerance to salinity, drought  
CC and temperature. The transgenic plants generated can be monocots or  
CC dicots and are especially maize, wheat, rice, oat, triticale, rice,  
CC barley, cotton, rapeseed, cassava, sunflower, tagetes, leguminous plants  
CC (e.g., soybean, peanut, alfalfa), solanaceous plants  
CC (e.g., potato, tobacco, aubergine, pepper, tomato), coffee, cacao, tea,  
CC Salix species, oil palm, coconut, perennial grasses and forage crops. The  
CC PKSRP nucleic acid and proteins may also be used in evolutionary and  
CC protein structural studies and as markers for specific regions of  
CC the genome.  
XX  
SQ Sequence 2177 BP; 522 A; 508 C; 547 G; 600 T; 0 other;  
Query Match 100.0%; Score 2177; DB 24; Length 2177;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 ggaagatcgaggttcacatggtcctacactcgtttctgtgtcagcgttcaaggtta 480  
DB 421 ggaagatcgaggttcacatggtcctacactcgtttctgtgtcagcgttcaaggtta 480  
QY 481 gttccgggtgtgttaccggcagcgcaagtcgagaaatctgagcaatctgttagcttga 540  
DB 481 gttccgggtgtgttaccggcagcgcaagtcgagaaatctgttagcttga 540  
QY 541 cgatcttcttcgtgttaactcagactactatgtctgcggttagcacccttacaactaac 600  
DB 541 cgatcttcttcgtgttaactcagactactatgtctgcggttagcacccttacaactaac 600  
QY 601 caatcgtatgacgagcggaaggtgtacagatgtgggtgtacactactcgtttgacgaac 660  
DB 601 caatcgtatgacgagcggaaggtgtacagatgtgggtgtacactactcgtttgacgaac 660  
QY 661 gccaaatgcccacacttcttagaagtcagcgcgaaggtagatcaatcgagatcccaagat 720  
DB 661 gccaaatgcccacacttcttagaagtcagcgcgaaggtagatcaatcgagatcccaagat 720  
QY 721 ttgcacagatttccaaacatggtgtccatgtgcggtgtgcgaagcccgacgacatgg 780  
DB 721 ttgcacagatttccaaacatggtgtccatgtgcggtgtgcgaagcccgacgacatgg 780  
QY 781 gagcacaacgtcacggaaggtttgtgtatctcatcagtaacgtacgtacgacgttgaac 840  
DB 781 gagcacaacgtcacggaaggtttgtgtatctcatcagtaacgtacgtacgacgttgaac 840  
QY 841 catcgaaagagagggccagaggtctccacgtcgtgtgagacgacatcgaaacactttaa 900  
DB 841 catcgaaagagagggccagaggtctccacgtcgtgtgagacgacatcgaaacactttaa 900  
QY 901 gctcgttttcaatcgtgacgttgcgtgcgagacgttcgactcctcagaattcaataagact 960  
DB 901 gctcgttttcaatcgtgacgttgcgtgcgagacgttcgactcctcagaattcaataagact 960  
QY 961 aatggaatcgccacactccttaactcttcggagagagccttatcgccgacggaactccaa 1020  
DB 961 aatggaatcgccacactccttaactcttcggagagagccttatcgccgacggaactccaa 1020  
QY 1021 agacgaattactcagatgatatataaagaatgcccgaaggttcgaaatagaatgtatct 1080  
DB 1021 agacgaattactcagatgatatataaagaatgcccgaaggttcgaaatagaatgtatct 1080  
QY 1081 ttcgaaagagacgaagaattctatcttcgaaagagcgtaagcggttaagtttgcgtgaa 1140  
DB 1081 ttcgaaagagacgaagaattctatcttcgaaagagcgtaagcggttaagtttgcgtgaa 1140  
QY 1141 gccggttttcgaagacatctcgaagatgaaagcctgcatagtaatccacactcgaatggc 1200  
DB 1141 gccggttttcgaagacatctcgaagatgaaagcctgcatagtaatccacactcgaatggc 1200  
QY 1201 ttaccggatagatggtgtctccggcgaagaagccttaccatcttaactctgaccgaa 1260  
DB 1201 ttaccggatagatggtgtctccggcgaagaagccttaccatcttaactctgaccgaa 1260  
QY 1261 gaacatttcaacgcgacagcaaggtgaggtgtcccgcaagagagcagctcctccgcgaaga 1320  
DB 1261 gaacatttcaacgcgacagcaaggtgaggtgtcccgcaagagagcagctcctccgcgaaga 1320  
QY 1321 cgtctcaacacactcccccagccgacatttgtagcggtgactggtgcgaagatctctgcgc 1380  
DB 1321 cgtctcaacacactcccccagccgacatttgtagcggtgactggtgcgaagatctctgcgc 1380  
QY 1381 atctcagcaagttcaacgaaatcgaggaacgcaaaatctcttatalatggtcagtttgt 1440  
DB 1381 atctcagcaagttcaacgaaatcgaggaacgcaaaatctcttatalatggtcagtttgt 1440  
QY 1441 gttcgcagactcattcattcattcccccatacagataaactcgtctgtgataaatt 1500  
DB 1441 gttcgcagactcattcattcattcccccatacagataaactcgtctgtgataaatt 1500



QY 1168 gaagcctcgcagatccacacccgagcttaccggaatagtggtctccgcga 1227  
|||||  
Db 310 gaagcctcgcagatccacacccgagcttaccggaatagtggtctccgcga 369  
|||||  
QY 1228 gaagcctcgcagatccacacccgagcttaccggaatagtggtctccgcga 1287  
|||||  
Db 370 gaagcctcgcagatccacacccgagcttaccggaatagtggtctccgcga 429  
|||||  
QY 1288 tctccgcgaagagcgagctctccgcgaagagcttaccaccccccgcg -ca 1345  
|||||  
Db 430 tgcgcgaagagcgagctctccgcgaagagcttaccaccccccgcg 489  
|||||  
QY 1346 tctgtagcgcgactgcg -tcgaatctctgcctcagcaggaagttcaagaaatcg 1404  
|||||  
Db 490 tctgtagcgcgactgcgactgcgaatctctgcctcagcaggaagttcaagaaatcg 549  
|||||  
QY 1405 a-ggcaagcgaaatctc-ttatacgctagtggtctc -cgaactcctcctatc 1461  
|||||  
Db 550 agggcaagcgaaatctccttatacgctagtggtctcctcgaactgcctatct 609  
|||||  
QY 1462 atcccccatcgagataactgcctctgctggaataattcctcaactt 1510  
|||||  
Db 610 at--cccatcgagataactg -atcgcgtggaataattcctcaacttt 655  
|||||

RESULT 3  
AAx84332/c  
ID AAx84332 standard; DNA: 5059 BP.  
XX  
AC AAx84332;  
XX  
DT 08-SEP-1999 (first entry)  
XX  
DE Stealth virus nucleic acid clone, SEQ ID NO: 24.  
XX  
KM Stealth virus; detection; diagnosis; infection; ss.  
XX  
OS Stealth virus.  
XX  
FH Key Location/Qualifiers  
FT misc\_difference 3605  
FT /\*tag- a  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3610  
FT /\*tag- b  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3615  
FT /\*tag- c  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3630  
FT /\*tag- d  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3631  
FT /\*tag- e  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3638  
FT /\*tag- f  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3641  
FT /\*tag- g  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"

FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3642  
FT /\*tag- h  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3657  
FT /\*tag- i  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3659  
FT /\*tag- j  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3681  
FT /\*tag- k  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3698  
FT /\*tag- l  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3703  
FT /\*tag- m  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3708  
FT /\*tag- n  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3713  
FT /\*tag- o  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3720  
FT /\*tag- p  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3732  
FT /\*tag- q  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3733  
FT /\*tag- r  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3740  
FT /\*tag- s  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3747  
FT /\*tag- t  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3750  
FT /\*tag- u  
FT /\*note- "this nucleotide is represented as a \* in the  
FT specification, and is included to maintain the  
FT base numbering given in the specification"  
FT  
FT misc\_difference 3751  
FT





PR	14-AUG-2000	2000S-0225268
PR	14-AUG-2000	2000S-0225270
PR	14-AUG-2000	2000S-0225271
PR	14-AUG-2000	2000S-0225277
PR	14-AUG-2000	2000S-0225278
PR	14-AUG-2000	2000S-0225279
PR	18-AUG-2000	2000S-0226579
PR	22-AUG-2000	2000S-0226681
PR	22-AUG-2000	2000S-0226688
PR	22-AUG-2000	2000S-0227182
PR	23-AUG-2000	2000S-0227180
PR	30-AUG-2000	2000S-0228994
PR	01-SEP-2000	2000S-0228287
PR	01-SEP-2000	2000S-0228345
PR	01-SEP-2000	2000S-0228344
PR	01-SEP-2000	2000S-0228354
PR	05-SEP-2000	2000S-0228509
PR	05-SEP-2000	2000S-0228513
PR	06-SEP-2000	2000S-0230437
PR	06-SEP-2000	2000S-0230438
PR	08-SEP-2000	2000S-0231242
PR	08-SEP-2000	2000S-0231244
PR	08-SEP-2000	2000S-0231413
PR	08-SEP-2000	2000S-0231414
PR	08-SEP-2000	2000S-0232080
PR	08-SEP-2000	2000S-0232081
PR	12-SEP-2000	2000S-0231968
PR	14-SEP-2000	2000S-0233397
PR	14-SEP-2000	2000S-0233399
PR	14-SEP-2000	2000S-0233399
PR	14-SEP-2000	2000S-0234001
PR	14-SEP-2000	2000S-0234240
PR	14-SEP-2000	2000S-0233063
PR	14-SEP-2000	2000S-0233064
PR	14-SEP-2000	2000S-0233065
PR	21-SEP-2000	2000S-0234223
PR	21-SEP-2000	2000S-0234274
PR	25-SEP-2000	2000S-0234997
PR	25-SEP-2000	2000S-0235484
PR	26-SEP-2000	2000S-0235484
PR	27-SEP-2000	2000S-0236327
PR	27-SEP-2000	2000S-0236327
PR	29-SEP-2000	2000S-0236358
PR	29-SEP-2000	2000S-0236358
PR	29-SEP-2000	2000S-0236359
PR	29-SEP-2000	2000S-0236360
PR	29-SEP-2000	2000S-0236362
PR	02-OCT-2000	2000S-0237037
PR	02-OCT-2000	2000S-0237038
PR	02-OCT-2000	2000S-0237040
PR	02-OCT-2000	2000S-0237043
PR	13-OCT-2000	2000S-0239937
PR	13-OCT-2000	2000S-0239937
PR	20-OCT-2000	2000S-0240960
PR	20-OCT-2000	2000S-0241785
PR	20-OCT-2000	2000S-0241786
PR	20-OCT-2000	2000S-0241807
PR	20-OCT-2000	2000S-0241808
PR	20-OCT-2000	2000S-0241809
PR	20-OCT-2000	2000S-0241826
PR	20-OCT-2000	2000S-0242221
PR	01-NOV-2000	2000S-0244617
PR	01-NOV-2000	2000S-0244617
PR	08-NOV-2000	2000S-0246475
PR	08-NOV-2000	2000S-0246475
PR	08-NOV-2000	2000S-0246476
PR	08-NOV-2000	2000S-0246477
PR	08-NOV-2000	2000S-0246478
PR	08-NOV-2000	2000S-0246524
PR	08-NOV-2000	2000S-0246524
PR	08-NOV-2000	2000S-0246525
PR	08-NOV-2000	2000S-0246526

XX	PR	08-NOV-2000;	2000US-0246527.
XX	PR	08-NOV-2000;	2000US-0246528.
XX	PR	08-NOV-2000;	2000US-0246532.
XX	PR	08-NOV-2000;	2000US-0246538.
XX	PR	08-NOV-2000;	2000US-0246609.
XX	PR	08-NOV-2000;	2000US-0246610.
XX	PR	08-NOV-2000;	2000US-0246611.
XX	PR	17-NOV-2000;	2000US-0249207.
XX	PR	17-NOV-2000;	2000US-0249208.
XX	PR	17-NOV-2000;	2000US-0249209.
XX	PR	17-NOV-2000;	2000US-0249210.
XX	PR	17-NOV-2000;	2000US-0249211.
XX	PR	17-NOV-2000;	2000US-0249212.
XX	PR	17-NOV-2000;	2000US-0249213.
XX	PR	17-NOV-2000;	2000US-0249214.
XX	PR	17-NOV-2000;	2000US-0249215.
XX	PR	17-NOV-2000;	2000US-0249216.
XX	PR	17-NOV-2000;	2000US-0249217.
XX	PR	17-NOV-2000;	2000US-0249218.
XX	PR	17-NOV-2000;	2000US-0249244.
XX	PR	17-NOV-2000;	2000US-0249245.
XX	PR	17-NOV-2000;	2000US-0249264.
XX	PR	17-NOV-2000;	2000US-0249265.
XX	PR	17-NOV-2000;	2000US-0249297.
XX	PR	17-NOV-2000;	2000US-0249299.
XX	PR	17-NOV-2000;	2000US-0249300.
XX	PR	01-DEC-2000;	2000US-0250391.
XX	PR	01-DEC-2000;	2000US-0251160.
XX	PR	05-DEC-2000;	2000US-0251030.
XX	PR	05-DEC-2000;	2000US-0251988.
XX	PR	05-DEC-2000;	2000US-0256729.
XX	PR	06-DEC-2000;	2000US-0251479.
XX	PR	08-DEC-2000;	2000US-0251856.
XX	PR	08-DEC-2000;	2000US-0251868.
XX	PR	08-DEC-2000;	2000US-0251869.
XX	PR	08-DEC-2000;	2000US-0251989.
XX	PR	08-DEC-2000;	2000US-0251990.
XX	PR	11-DEC-2000;	2000US-0254097.
XX	PR	05-JAN-2001;	2001US-0259678.
FA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Rosen CA, Barash SC, Ruben SM;		
DR	WPI; 2001-541565/60.		
PT	Nucleic acids encoding 3224 human nervous system antigen polypeptides,		
PT	useful for preventing, diagnosing and/or treating nervous system		
PT	cancers and metastases -		
XX	Disclosure; SEQ ID NO 7348; 1701pp + Sequence Listing; English.		
XX	The invention relates to novel genes (ABA11004-ABA21534) and proteins		
CC	(ABA11678-ABA18001) useful for preventing, treating or ameliorating		
CC	medical conditions e.g. by protein or gene therapy. The genes are		
CC	isolated from a range of human tissues disclosed in the specification.		
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful		
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast		
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone		
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;		
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune		
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's		
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative		
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;		
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and		
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal		
CC	and parasitic infections.		
CC	Note: The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pat_sequences.		
XX	Sequence 2625 BP; 699 A; 644 C; 581 G; 701 T; 0 other:		













CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SQ Sequence 8375 BP; 2190 A; 134 C; 2005 G; 4046 T; 0 other;

Query Match 1.5%; Score 33.4; DB 24; Length 8375;  
 Best Local Similarity 49.7%; Pred. No. 17;  
 Matches 85; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
 QY 1977 taccatcgactagcgtcactttgactcgttgatataatcgtttgcaattagtgagactt 2036  
 || || || | | | | | | | | | | | | | | | |  
 Db 1710 tagaatgagttaaaaaaatttaaaaatggtttttgtttgtttttgtttgtttt 1769  
 QY 2037 tctaatgatagatgacatttgacagtacgttgatgctgtttacaagatagcgagc 2096  
 | |||| | | | | | | | | | | | | | | | |  
 Db 1770 ttgagatggaatttctgtttctgttttaggttgagtcgtgagtcgacgatttcgttga 1829  
 QY 2097 tagaagccttaaacctttagctacacatgtatatttaaacctatagagt 2147  
 | | | | | | | | | | | | | | | | | | | |  
 Db 1830 tcgtaatttcgltttttaagtttaagtaattttttttaagttttaagt 1880

Search completed: September 15, 2002, 23:46:06  
 Job time: 5230 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2002, 21:47:06 ; Search time 59.1 Seconds  
(without alignments)  
9048.124 Million cell updates/sec

Title: US-09-828-313-22

Perfect score: 2177  
Sequence: 1 atcccgagctgtatgtgct.....gcagatattgcgttaacgc 2177

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	41.6	1.9	7218	1 US-08-232-463-14	Sequence 14, Appl
2	31.8	1.5	1935	2 US-08-878-989-11	Sequence 11, Appl
3	31.8	1.5	1935	4 US-09-272-796-11	Sequence 11, Appl
4	31.8	1.5	4403765	4 US-09-103-840A-2	Sequence 2, Appl
5	31.8	1.5	4411529	4 US-09-103-840A-1	Sequence 1, Appl
6	31.2	1.4	396	3 US-08-963-168C-4	Sequence 4, Appl
7	31.2	1.4	458	3 US-09-141-000-4	Sequence 4, Appl
8	30.4	1.4	1101	4 US-09-078-294-20	Sequence 20, Appl
9	30.4	1.4	1791	1 US-08-245-294-7	Sequence 7, Appl
10	30.4	1.4	1791	1 US-08-474-499-7	Sequence 7, Appl
11	30.4	1.4	1791	1 US-08-307-279A-7	Sequence 7, Appl
12	30.4	1.4	1791	5 PCT-US95-06211-7	Sequence 7, Appl
13	30.4	1.4	4164	1 US-08-204-675-1	Sequence 1, Appl
14	30.4	1.4	4164	1 US-08-660-754-1	Sequence 1, Appl
15	30.4	1.4	4164	2 US-08-796-364-1	Sequence 1, Appl
16	30.4	1.4	4164	5 PCT-US95-02520-1	Sequence 1, Appl
17	30.4	1.4	4933	1 US-08-204-675-2	Sequence 2, Appl
18	30.4	1.4	4933	2 US-08-660-754-2	Sequence 2, Appl
19	30.4	1.4	4933	2 US-08-796-364-2	Sequence 2, Appl
20	30.4	1.4	4933	5 PCT-US95-02520-2	Sequence 2, Appl
21	30.4	1.4	5181	3 US-08-801-344-5	Sequence 5, Appl
22	30.4	1.4	5181	3 US-09-438-599-5	Sequence 5, Appl
23	30.4	1.4	7102	3 US-09-138-024-20	Sequence 20, Appl
24	30.4	1.4	7333	3 US-09-138-024-21	Sequence 21, Appl
25	30.4	1.4	7633	3 US-09-028-851-1	Sequence 1, Appl
26	30.4	1.4	7633	3 US-08-815-520-1	Sequence 1, Appl
27	30.4	1.4	7633	4 US-09-273-163-1	Sequence 1, Appl

28	30.4	1.4	9936	4 US-08-972-927-2	Sequence 2, Appl
29	30	1.4	2933	4 US-08-936-165A-201	Sequence 201, App
30	30	1.4	5163	3 US-08-700-651-1	Sequence 1, Appl
31	30	1.4	5163	3 US-08-928-361B-4	Sequence 4, Appl
32	30	1.4	5318	3 US-08-700-651-2	Sequence 2, Appl
33	30	1.4	5318	3 US-08-928-361B-3	Sequence 3, Appl
34	30	1.4	5511	3 US-08-928-361B-2	Sequence 2, Appl
35	30	1.4	7334	3 US-08-928-361B-1	Sequence 1, Appl
36	29.8	1.4	1342	3 US-08-445-515-54	Sequence 54, Appl
37	29.8	1.4	1558	1 US-08-467-607-2	Sequence 2, Appl
38	29.8	1.4	1558	2 US-08-469-362-2	Sequence 2, Appl
39	29.8	1.4	1558	2 US-08-850-392-2	Sequence 2, Appl
40	29.8	1.4	6253	3 US-08-893-327-15	Sequence 15, Appl
41	29.8	1.4	6280	3 US-08-893-327-17	Sequence 17, Appl
42	29.8	1.4	6280	3 US-08-893-327-19	Sequence 19, Appl
43	29.6	1.4	2166	2 US-08-317-401E-3	Sequence 3, Appl
44	29.6	1.4	2877	2 US-08-317-401E-1	Sequence 1, Appl
45	29.6	1.4	3591	1 US-07-943-843-5	Sequence 5, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
: Sequence 14, Application US/08232463  
: Patent No. 5670367  
: GENERAL INFORMATION:  
: APPLICANT: DORNER, F.  
: APPLICANT: SCHEIFLINGER, F.  
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
: NUMBER OF SEQUENCES: 52  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: Foley & Lardner  
: STREET: 1800 Diagonal Road, Suite 500  
: CITY: Alexandria  
: STATE: VA  
: COUNTRY: USA  
: ZIP: 22313-0299  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/232,463  
: FILING DATE:  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US/07/935,313  
: FILING DATE:  
: APPLICATION NUMBER: EP 91 114 300.6  
: FILING DATE: 26-AUG-1991  
: ATTORNEY/AGENT INFORMATION:  
: NAME: BENT, Stephen A.  
: REGISTRATION NUMBER: 29,768  
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (703)836-9300  
: TELEFAX: (703)683-4109  
: TELEX: 899149  
: INFORMATION FOR SEQ ID NO: 14:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 7218 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: IMMEDIATE SOURCE:  
: CLONE: PTZgpt-F15  
: US-08-232-463-14





STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SYNORAT04  
CLONE: 70365  
US-09-272-796-11

Query Match 1.5%; Score 31.8; DB 4; Length 1935;  
Best Local Similarity 51.8%; Pred. No. 2.7;  
Matches 72; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1489 ttggataattctccaaactlttgctcttcattccctcaagaagctctcaatgagcagt 1548  
DB 1593 ttgagaaatccctaaatcctgtagaatttctctgtagtcataatgacacacattcttcagg 1534  
QY 1549 aataagttacgacatttgacaaacttaagcagtgatttctgtatcccaagcttc 1608  
DB 1533 acgatttcagctgattgttatctaccccaagacacctggtgattcagctgaggttc 1474

QY 1609 atcgaggtatcaaggaatg 1627  
DB 1473 tctgtgtatggtgtaatg 1455

RESULT 4  
US-09-103-840A-2/c  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
US-09-103-840A-2

Query Match 1.5%; Score 31.8; DB 4; Length 4403765;  
Best Local Similarity 52.7%; Pred. No. 54;  
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 817 agtaacgagtcagcgctgataccatcacgaagggcccaagcttaacctctg 876  
DB 4327020 actgcttgacacgacgactggtccgctcgcagattccggagggcaaaagacgtccga 4326961  
QY 877 agacgacatcagaagacaccttaagctgttccatctggcattgtctgagacactcg 936  
DB 4326960 tgccccaacatcgaatcgtatgctgacgagttgcatcgctatctacgagagctca 4326901  
QY 937 acttctagaa 947  
DB 4326900 acctgctgcta 4326890

RESULT 5  
US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
Patent No. 6294328

GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 1.5%; Score 31.8; DB 4; Length 4411529;  
Best Local Similarity 52.7%; Pred. No. 54;  
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 817 agtaacgagtcagcgctgataccatcacgaagggcccaagcttaacctctg 876  
DB 4334766 agtggcttgacacgacgactggtccgctcgcagattgcccgaagacacgtccga 4334707  
QY 877 agaccgacatcgaagacaccttaagctgttccatcgcgacattgtctgagacactcg 936  
DB 4334706 tgccccaacatcgaatcgtatgctgacgagttgcatcgctatctacgagagctca 4334647  
QY 937 acttctagaa 947  
DB 4334646 acctgctgcta 4334636

RESULT 6  
US-08-963-168C-4/c  
Sequence 4, Application US/08963168C  
Patent No. 6127166  
GENERAL INFORMATION:  
APPLICANT: Bayley, Hagan  
APPLICANT: Cao, Quiding  
APPLICANT: Wang, Yunjuan  
TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES  
TITLE OF INVENTION: AND GENES ENCODING THEM  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963,168C  
FILING DATE: 03-NOV-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, Peter J.  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07917/059001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 396 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...393  
US-08-963-168C-4

Query Match 1.4%; Score 31.2; DB 3; Length 396;  
Best Local Similarity 45.8%; Pred. No. 1.5;  
Matches 108; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Oy 1120 gcgagcgaagttgctgaggaagcggtttgcagagatctccggaagaagcctgcat 1179  
Db 339 GCCCCTCCCATTCGCGCGAAMCGGCAFTTCCGCTCCCATTCCTCGAAMCCTCCCAT 280  
Oy 1180 ggaattccacacatcgatggcttaaccggatagtggtctcgcggaagacgctctac 1239  
Db 279 ACCACTGCCCATTCCTCGGAATCCACTTTCGCGCGCATTCGCGGAAACCTCCCATTC 220  
Oy 1240 cccatctaccctgcggcggaagaacatttcacgcgcgaacgaagtgatggtccgcaag 1299  
Db 219 ACCGAATCCGCTTGACCTCCACCATACCACTCCGAAAGCGGCTATTCCGCGGAATCC 160  
Oy 1300 gacgagctcctcgcgcgaagaacgtctcacacctcccgccgacgttgtagcgc 1355  
Db 159 GCCTTTACTCTCCGCGCATTCGCGCATACCGCGAATCCGCTTGCTTCATTCGCGC 104

## RESULT 7

US-09-141-000-4/C  
Sequence 4, Application US/09141000  
Patent No. 6034295  
GENERAL INFORMATION:  
APPLICANT: Chen, Fang  
TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NUCLEAR  
TITLE OF INVENTION: RECEPTOR PROTEINS  
FILE REFERENCE: 199999Y  
CURRENT APPLICATION NUMBER: US/09/141.000  
CURRENT FILING DATE: 1998-08-26  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 458  
TYPE: PRT  
ORGANISM: Human  
US-09-141-000-4

Query Match 1.4%; Score 31.2; DB 3; Length 458;  
Best Local Similarity 8.7%; Pred. No. 1.6;  
Matches 34; Conservative 121; Mismatches 237; Indels 0; Gaps 0;

Oy 1419 cctctataggggtatgttgctgcagacgcacccctatctatcccccataagata 1478  
Db 442 C.M.NR.D.BTMAS.Y...AK.KMCTTY.H.KD.CT.RH..T.D.BH..M.BT.BH. 383  
Oy 1479 actgcatctggtgataaattctccaacatttgctctcctcgaagacgctctc 1538  
Db 382 .DKSHSNT.T.TW.AB...M.MKSMRMB...TNN.H..CT.MS.H.HK.RHHTRB.. 323  
Oy 1539 aatggcagaaatagtgtagacattgacacacatccaattagctatgctatctcgta 1598  
Db 322 H..S.STRBBC...KMTS...SK.HT.S.AS.C..DMTWC..BB..YHT.HG.AA.TM 263  
Oy 1599 acccagctcatcgagatgatacaagaaatgagcgacgaagaacactgctactgtctgg 1658  
Db 262 .HSH.BA.H.KTR.M...TB...DSB.MNR.MMT..B...N.R.SN.TH.YWMRM.YCCYB 203  
Oy 1659 tatccgctgtgacgagatgcatgctgacacgctgctatcagtgagatttcttgagcg 1718

Db 202 YH.YBC.M.KCBM.GMK.YGT.GSMYYMA..G.NAT.GSR..NC..AYM..TMGT.GST 143  
Oy 1719 cagatctgctccgcagattgtgttataacgtttgctgtagagggcctagacgta 1778  
Db 142 BCRDRGST.HCGB.G.YM.KSN..KR.GMA..H.B.ASSGHRH.MVB.CSCC...TS.R.. 83  
Oy 1779 ctatcaagcaatgagaagtgtgctgtgtgga 1810  
Db 82 .SH.CN.DCNKASSRSCSTHSCC.SDDNBH 51

## RESULT 8

US-09-078-294-20  
Sequence 20, Application US/09078294  
Patent No. 6265211  
GENERAL INFORMATION:  
APPLICANT: Choo, Kong-Hong Andy  
APPLICANT: Du Sart, Desiree  
APPLICANT: Cancilla, Michael R.  
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
FILE REFERENCE: Davies Col  
CURRENT APPLICATION NUMBER: US/09/078.294  
CURRENT FILING DATE: 1998-05-13  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 20  
LENGTH: 1101  
TYPE: DNA  
ORGANISM: BAC-F2 contig 47  
US-09-078-294-20

Query Match 1.4%; Score 30.4; DB 4; Length 1101;  
Best Local Similarity 55.8%; Pred. No. 5.6;  
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Oy 2064 tagcttgatcgtcttcaagaatagcgacgtagaagccttaaccttaacctaccat 2123  
Db 412 tatctgcatgatttctccatgagggacagtaagcgttaagcgttaagcattatccgcaa 471  
Oy 2124 gatatattaaacctatagaatgaaagcgctgtagatattg 2167  
Db 472 gtacaatttctactcttcgaagacagaataattgctgacatg 515

## RESULT 9

US-08-245-294-7/C  
Sequence 7, Application US/08245294  
Patent No. 5644047  
GENERAL INFORMATION:  
APPLICANT: Anderson, Burt E.  
APPLICANT: Regnery, Russell L.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
TITLE OF INVENTION: DIAGNOSING  
TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: NEEDLE & ROSENBERG,P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,294  
FILING DATE:

CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414.612  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 141..1649  
US-08-245-294-7

Query Match 1.4%; Score 30.4; DB 1; Length 1791;  
Best Local Similarity 50.0%; Pred. No. 7.8;  
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1459 tctattcccatcgagataactgcatcggttgataaattctccaacatttgcctc 1518  
DB 1390 tccaatgtttatctgaatccgctcatttgaaatatttgagccattcttcatattt 1331  
QY 1519 tcattcctaagcagctcctcaatgagccagtaataatgtaacacattgacaaactccaa 1578  
DB 1330 tgcgtcttcagcagcagatcgaatcgaatttaacatttcttctttaccacatttccaa 1271  
QY 1579 ttaagtaagcttattcgttaacccaagttcat 1610  
DB 1270 actcctaaagttactgtttctcctgggctcat 1239

RESULT 10  
US-08-474-499-7/C  
Sequence 7, Application US/08474499  
Patent No. 5693776  
GENERAL INFORMATION:  
APPLICANT: Anderson, Burt E.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE  
TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474.499  
CLASSIFICATION: 536  
FILING DATE: 07-JUN-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/245,294  
FILING DATE: 18-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414.612  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 141..1649  
US-08-474-499-7

Query Match 1.4%; Score 30.4; DB 1; Length 1791;  
Best Local Similarity 50.0%; Pred. No. 7.8;  
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1459 tctattcccatcgagataactgcatcggttgataaattctccaacatttgcctc 1518  
DB 1390 tccaatgtttatctgaatccgctcatttgaaatatttgagccattcttcatattt 1331  
QY 1519 tcattcctaagcagctcctcaatgagccagtaataatgtaacacattgacaaactccaa 1578  
DB 1330 tgcgtcttcagcagcagatcgaatcgaatttaacatttcttctttaccacatttccaa 1271  
QY 1579 ttaagtaagcttattcgttaacccaagttcat 1610  
DB 1270 actcctaaagttactgtttctcctgggctcat 1239

RESULT 11  
US-08-307-279A-7/C  
Sequence 7, Application US/08307279A  
Patent No. 5736347  
GENERAL INFORMATION:  
APPLICANT: Anderson, Burt E.  
TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae  
TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea  
TITLE OF INVENTION: Henselae and Rochalimaea Quintana Infection  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: 127 Peachtree Street, N.E., Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/307,279A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414.624  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 688-0770  
TELEFAX: (404) 688-9880  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

FEATURE:  
NAME/KEY: COS  
LOCATION: 141..1652  
US-08-307-279A-7

Query Match 1.4%; Score 30.4; DB 1; Length 1791;  
Best Local Similarity 50.0%; Pred. No. 7.8;  
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1459 tctattcccccagagagatactgcttgtagataaattctccaaacatttgcctc 1518  
DB 1390 tccaatgttttcatcgaattacccgctcatattgaaattttgaccacattctcatattt 1331  
QY 1519 tcatcccaagacgcccctccaaatgagtaataatgttaagacattgtgcacaaactccaa 1578  
DB 1330 tgcgtcttcagacatcgaattacggttaattatattcttctttacacgaatttcgaa 1271  
QY 1579 ttaagtagcttattctgttaaccacgctcat 1610  
DB 1270 actcctaaggttactgttctctcgctcat 1239

RESULT 12  
PCT-US95-06211-7/c  
Sequence 7, Application PC/TUS9506211  
GENERAL INFORMATION:

APPLICANT:  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
TITLE OF INVENTION: ROCHALLIMAEA HENSELAE AND ROCHALLIMAEA QUINTANA INFECTION  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P. C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06211

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/245,294  
FILING DATE: 18 MAY 1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Spratl, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414.6121  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

MOLECULE TYPE: DNA (genomic)  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: COS  
LOCATION: 141..1649

Query Match 1.4%; Score 30.4; DB 5; Length 1791;  
Best Local Similarity 50.0%; Pred. No. 7.8;  
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1459 tctattcccccagagagatactgcttgtagataaattctccaaacatttgcctc 1518  
DB 1390 tccaatgttttcatcgaattacccgctcatattgaaattttgaccacattctcatattt 1331  
QY 1519 tcatcccaagacgcccctccaaatgagtaataatgttaagacattgtgcacaaactccaa 1578  
DB 1330 tgcgtcttcagacatcgaattacggttaattatattcttctttacacgaatttcgaa 1271  
QY 1579 ttaagtagcttattctgttaaccacgctcat 1610  
DB 1270 actcctaaggttactgttctctcgctcat 1239

RESULT 13  
US-08-204-675-1  
Sequence 1, Application US/08204675  
Patent No. 5677170  
GENERAL INFORMATION:

APPLICANT: Devine, Scott E.  
APPLICANT: Boeke, Jef D.

APPLICANT: Braiterman, Lelita T.  
TITLE OF INVENTION: In Vitro Transposition of Artificial  
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie, and Beckett  
STREET: 1001 G Street, N.W.

CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.

ZIP: 20001

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/204,675  
FILING DATE: 02-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.45501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202.508.9100  
TELEFAX: 202.508.9299  
TELEX: 197430 BBMB UR

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4164 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: PAT-1  
US-08-204-675-1

Query Match 1.4%; Score 30.4; DB 1; Length 4164;  
Best Local Similarity 55.8%; Pred. No. 14;  
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 2064 tagcttgagatcgtttacagatagcgcagcttaagccttaaaccttagctacat 2123  
DB 605 tatcttgacatgatttttccatgagcgcacggttaagccgcttaagacgattatccgcaa 664  
QY 2124 gattattaaacatataagcgaagcgcgttgagatattg 2167  
DB 665 gtacaaatttttactcttcgaagacagaaattttgctgacattg 708



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2002, 21:28:31 : Search time 1744.24 Seconds  
(without alignments)  
16845.658 Million cell updates/sec

Title: US-09-828-313-22

Sequence: 2177 1 atcccgagcttgatgtgct.....gcagatattgcgttaacgc 2177

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_estl:\*
- 10: gb\_estl2:\*
- 11: gb\_hic:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_hiv:\*
- 15: em\_gss\_lnv:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	290.8	13.4	294	10	BUI179058	BUI179058 BUI179058
2	202.8	9.3	547	10	BUI174514	BUI174514 BUI174514
3	113.6	5.2	605	10	BUI174924	BUI174924 BUI174924
4	50.4	2.3	526	9	AW739415	AW739415 gba1f06.y
5	42	1.9	855	12	CNS011PE	AL100556 Drosophila
6	39.8	1.8	949	12	AG029463	AG029463 Pan trogl
7	39.2	1.8	1101	12	CNS00100	AL068607 Drosophila
8	38.6	1.8	553	9	AM625386	AM625386 EST319209
9	38.6	1.8	671	9	AW930530	AW930530 Pan trogl
10	37.8	1.7	577	12	AG155394	AG155394 Pan trogl
11	37.4	1.7	513	10	BG367216	BG367216 HVSME000
12	37.4	1.7	940	10	BF623571	BF623571 HVSME000
13	37.4	1.7	1001	12	CNS00HG7	AL073444 Drosophila
14	37.2	1.7	513	12	B87304	B87304 RPL11-28H1
15	37.2	1.7	1067	12	CNS03PF9	AL241290 Tetraodon
16	37.2	1.7	507	9	A1775253	A1775253 EST256353
17	37	1.7	780	10	BM407305	BM407305 EST581632

18	36.8	1.7	501	10	BG405233	BG405233 saca9h03.
19	36.8	1.7	872	10	BF159188	BF159188 60176634
20	36.8	1.7	1028	12	CNS00JPV	AL076788 Drosophila
21	36.6	1.7	497	12	A0772166	A0772166 HS_5432.B
22	36.6	1.7	611	9	BE111776	BE111776 UT-R-B1-
23	36.6	1.7	952	9	BR037533	BR037533 MPT1G02.M
24	36.4	1.7	307	9	BB104250	BB104250 BB104250
25	36.4	1.7	1067	10	BF529057	BF529057 602040054
26	36.4	1.7	1101	12	CNS000D1	AL065414 Drosophila
27	36.2	1.7	438	10	BM136557	BM136557 WHEE213.C
28	36.2	1.7	449	12	AZ498999	AZ498999 1M033603
29	36.2	1.7	507	12	AZ498997	AZ498997 1M033603
30	36.2	1.7	1101	12	CNS0129R	AL101288 Drosophila
31	36.2	1.7	418	12	AO340838	AO340838 HS_5022.B
32	36	1.7	432	9	AT10421	AT10421 EST238714
33	36	1.7	469	10	BG726049	BG726049 sae07d03.
34	35.8	1.6	353	9	AT362949	AT362949 gy87f05.x
35	35.8	1.6	359	10	H62450	H62450 YF79f08.r1
36	35.8	1.6	836	10	BF623428	BF623428 HVSME000
37	35.6	1.6	286	9	BB415825	BB415825 BB415825
38	35.6	1.6	789	12	BH450439	BH450439 BOCG014TF
39	35.6	1.6	790	12	BH250073	BH250073 BOCAF37TF
40	35.4	1.6	415	12	AZ223450	AZ223450 RPL1-23-5
41	35.4	1.6	711	12	AZ319283	AZ319283 1M0038L16
42	35.4	1.6	778	9	AV384757	AV384757 AV384757
43	35.2	1.6	506	10	N95351	N95351 zb62909.s1
44	35.2	1.6	518	10	BE805900	BE805900 ss62910.Y
45	35.2	1.6	519	9	AT164309	AT164309 A059P10U

ALIGNMENTS

RESULT 1  
LOCUS BUI179058 294 bp mRNA linear EST 24-JAN-2002  
DEFINITION BUI179058 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens  
CDNA clone pphb24a07 5', mRNA sequence.  
BUI179058  
BUI179058.1 GI:18347012

ACCESSION  
VERSION BUI179058  
KEYWORDS  
SOURCE  
ORGANISM Physcomitrella patens subsp. patens.  
Physcomitrella patens subsp. patens  
Eukaryota; Viridiplantae; Streptophyta; Bryophyta;  
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

REFERENCE  
AUTHORS Fujita,T., Shin-I,T., Seki,M., Kamiya,A., Uchiyama,T., Nishiyama,T., Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe

TITLE  
JOURNAL  
COMMENT  
Comparison of the moss Physcomitrella patens genome with flowering plants genome  
Unpublished (2002)  
Contact: Tadasu Shin-I  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855

FEATURES  
source  
1..294  
Email: tsuh1@genes.nig.ac.jp  
A backbone of the vector is pBluescript II, that was in vivo excised from a modified lps phage vector (Mo bi Tec, Germany). XhoI digested-5' end of cDNA is ligated to SalI site of the vector, and the BamHI digested-3' end including poly-A tail is ligated to BamHI site of the vector. cDNA insert could be amplified with conventional T7 and T3 primers. This normalized full-length cDNA library was generated basically according to the method described in Genome Research 10, 1617-1630 (2000). Carninci, P. et al. Protonemata were blended for the POLYTRON, and then cultivated on the BCD medium containing 0.5uM BA (benzylaminopurine) for 8 to 13 days under the continuous light.  
Location/Qualifiers  
1..294





JOURNAL  
COMMENT

Plants genome  
Unpublished (2002)  
Contact: Tadasu Shin-1  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855

A backbone of the vector is plasmid vector II, that was *in vivo* excised from a modified I<sup>PS</sup> phase vector (Mo bi Tec, Germany). Shortly digested 5'-end of cDNA is ligated to SalI site of the vector, and the BamHI digested 3'-end including poly-A tail is ligated to BamHI site of the vector. cDNA insert could be amplified with conventional 7' and 3' primers. This normalized full-length cDNA library was generated basically according to the method described in genome Research 10, 1617-1650 (2000), Carriello, P. et al. Protozoanmata were blended by the POLYTRON, and then cultivated on the BCD medium containing 0.5mM BA (benzyladeninopurine) for 8 to 13 days under the continuous light.

## FEATURES

```

source
1. .605
/organism="Physcomitrella patens subsp. patens"
/db_xref="taxon:145481"
/clone="phb13n04"
/clone_1lb="normalized full length cDNA library,
chloronemata, caulonemata and malformed buds"
/tissue_type="mixture of chloronemata, caulonemata and
malformed buds"
BASE COUNT      141 a      125 c      145 g      194 t
ORIGIN

```

Query Match	5.28;	Score 113.6;	DB 10;	Length 605;
Best Local Similarity	73.68;	Pred. No. 3.9e-23;		
Matches 159;	Conservative 0;	Mismatches 54;	Indels 3;	Gaps 1

QY	440	atgggttcacatccggtttctttctgtgaacg--gttcaaggttctgacgggtttgtctac	456
Db	390	ATGGGGGCGATCATCTCTTTTCTTGGCATTTAGCCCGAGAGATGATTTGCTGCGGGTGTATTAT	448
QY	497	ccgaagcgcaagtcgagaaatctgaagcaatctctgtagcgtgcagatctcttcgtgt	556
Db	450	CCTGACGGTCACTGACGACAATTTGAGCAGAACTTGCAGACGCTGATAGCGTGGCTCTAGCT	509
QY	557	aatccacatcactatgctctgcgttagcagcccttacaacatcaccatcgtatgtgacgc	616
Db	510	AATCCAAATTACTTTGTTTGGGGGTGCACCCGTCACACTTACGAAATGATGATGACGAA	569
QY	617	gaagaggtgctgagatagagggtgactacttcgtt	652
Db	570	GATGAGCACTTGTGAAAGAGATTACCTACTTCGTT	605

RESULT	4
LOCUS	AW739415
DEFINITION	526 bp mRNA linear EST 25-APR-2000
ACCESSION	gb181f06.y1 Moss EST library PPN Physcomitrella patens cDNA clone
VERSION	ppr SOURCE_ID:PPN161512.5 , mRNA sequence.
KEYWORDS	AW739415.1 GI:7648632
SOURCE	EST.
ORGANISM	Physcomitrella patens.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
AUTHORS	Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
	1 (bases 1 to 526)
	Quatrano,R., Bashirades,S., Cove,D., Cuning,A., Knight,C., Clifton
	S., Marra,M., Hillier,L., Page,D., Martin,J., Wylie,T., Underwood
	K.T., Theising,B., Allen,M., Bowers,Y., Person,B., Sailer,T.,
	Steele,M., Gibbons,M., Harvey,N., Ritzer,E., Jackson,Y., McCann,R.,
	Waterston,R. and Wilson,R.
TITLE	leeds/Wash u Moss EST project

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Ralph Quatramo  
Leeds/Wash U Moss EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Libraries were constructed by Dr. Stavros Bashiardes as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)  
Seq. primer: -40RP from Glbco  
High quality sequence stop: 413.

FEATURES  
source

```

/organism="Physcomitrella patens"
/db_xref="taxon:3218"
/clone="PEP_SOURCE_ID:PPN161512"
/clone.lib="Moss EST library PPN"
/tissue.type="protonemata: 7 day old tissue auxin treated"
/lab.host="DH10B"
/notes="vector: plasmid SK-; Site.1: EcoRI. Site.2:
XhoI. Construction of the cDNA library was carried out
using Stratagene's 'UnizAP - cDNA synthesis kit'. cDNA was
constructed using an oligo dt primer/linker that contains
a XhoI site within it. Following ds cDNA synthesis,
EcoRI adapters were ligated to the blunt ends and sample
was digested with XhoI. The result is cDNA with an EcoRI
sticky end on one side and a XhoI sticky end on the other.
This cDNA was ligated directionally in UnizAP arms. The
vector is designed containing the plasmid sequence as
well as lambda DNA and cDNA is cloned within this
plasmid sequence. The vector was then packaged using
Gold digipackaging extracts. Library was grown in XLBlue
MR+ cells and amplified. The library was excised by mass
exaction using Stratagene's 'Mass excision kit' that uses
exaxist as a helper phage that releases the plasmid
sequence and circularises it as single stranded plasmids
that are then packaged (by helper phage) and secreted out
of the host cell as phagemids. SOLR cells were transformed
with phagemids and the library was plated out on LB-amp
plates to select for transformants. Approximately 1,000
,000 colonies were grown and recovered. The double
stranded plasmid library was recovered by using Unigen
Midi prep kit. 2 micro grams of each library were used to
transform DH10B cells by electroporation."

```

Query Match	2.3%;	Score 50.4;	DB 9;	Length 526;
Best Local Similarity	66.7%;	Pred. No. 0.00075;		
Matches	72;	Conservative	0;	Mismatches 36;
			Indels	0;
			Gaps	0;

[illegible]

KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 855)  
GENOSCOPE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (CEGP) -  
http://www.cegp.ebi.ac.uk . This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billard at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

FEATURES  
source Location/Qualifiers  
1..855  
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BAC06M20"  
/note="end : Sep6"  
BASE COUNT 96 a 63 c 75 g 160 t 461 others  
ORIGIN

Query Match 1.98; Score 42; DB 12; Length 855;  
Best Local Similarity 12.18; Pred. No. 0.4; Mismatches 344; Indels 0; Gaps 0;  
Matches 65; Conservative 129; Mismatches 344; Indels 0; Gaps 0;

1522 tctctcaagaagctctcctcaatgagcagatatagtttagacatgtgtgcaactccaatta 1581  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 694 KBKKMAACMCKNNKKKKKKKKKKKKNNNNNNNNNNNNNNNNNNNNNNNNNNNN 635  
1582 cgttagcgtatctctgtaaccagtcagtgaggtatcaaggaatgcgcgtagtaagact 1641  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 634 NNNKKKKNN 575  
1642 gctactctgtctgtgataccgctgtgacgagatgcatgtcgacacgtgctcatcag 1701  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 574 NNN 515  
1702 tgggaattcttgagcgcagatctgtcttcgcagattgttccaaagcttgtgtcgt 1761  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 514 KNNKKKKNN 455  
1762 agggagcctagaagcttactatcaagcaatgagaagctgtgtgtgtgtgtgtgtgtgt 1821  
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 454 NKGCGKKNNKKKKNN 395  
1822 tctttggaggttcttctctctatgtagaactatgagagacacttgcctgtgtgtgc 1881  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 394 NNNKKKKGGANNGKNN 335  
1882 acatccatagaacatagctcactctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1941  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 334 NNN 275  
1942 tctcatctatttgggt 2001  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 274 KKKKKKKTTGTGTGKTKKKKKKKKKKKKKGGAGKGAANNNNNNNTNNANNANTNANGN 215  
2002 tgcgtgataatcgttgcataattagtggaagctttagatagatagatgtcatttgt 2059  
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 214 TGAAGGGTATTGANGAAMAAGKGNATTTGTTTARAGSAGATTGTATATATTTT 157

RESULT 6  
LOCUS AG029463/c 949 bp DNA linear GSS 01-NOV-2001  
DEFINITION Pan troglodytes DNA, clone: PTB-001122.F, genomic survey sequence.  
ACCESSION AG029463  
VERSION AG029463.1 GI:16556336  
KEYWORDS GSS; GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male  
BAC library clone:PTB-001122.F.  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
REFERENCE 1 (sites)  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
TITLE Tozoki,Y., Watanabe,H. and Sakaki,Y.  
JOURNAL BAC end sequences of library PTB  
UNPUBLISHED  
2 (bases 1 to 949)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Tozoki,Y., Watanabe,H. and Sakaki,Y.  
DIRECT SUBMISSION  
TITLE Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimpdsgsc.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the Rad process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : PK5145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
LOCATION/QUALIFIERS  
1..949  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-001122.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC library"  
BASE COUNT 325 a 311 c 100 g 208 t 5 others  
ORIGIN

Query Match 1.88; Score 39.8; DB 12; Length 949;  
Best Local Similarity 59.18; Pred. No. 2; Mismatches 47; Indels 0; Gaps 0;  
Matches 68; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

148 tgtgatttgagggtcttgactgtagaagagctcttgagctgagaagtgtgcaa 207  
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 494 TGTGATTCGGCGCTGCTGTAAATTGAGAAATGCTGTGCTGGCGGCTGTTTG 435  
OY 208 tttagggcaccacgggt 262  
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 434 TTTAGGCGGAATTGATTGTGCGAGCTGTGCTTAGTCAATTGTTGCT 380

RESULT 7  
LOCUS CNS00100/c 1101 bp DNA linear GSS 03-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR32D23 of RPCI-98 library from Drosophila melanogaster (fruit  
fly) genomic survey sequence.  
ACCESSION ADB6607  
VERSION ADB6607.1 GI:4958669  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;







REFERENCE	ORGANISM	SOURCE
1 (bases 1 to 1001)	fruit fly.	
Genoscope.	Drosophila melanogaster	
Direct Submission		
Submitted (02-JUN-1999)	Genoscope - Centre National de Sequencage	
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
- Web : www.genoscope.cns.fr)		
Determination of this BAC-end sequence was carried out as part of a		
collaboration with the Berkeley Drosophila Genome Project (BDGP).		
The BDGP is constructing a physical map of the Drosophila		
melanogaster genome using these BACs. For further information		
please see <a href="http://www.fruitfly.org/TheBDGP/Drosophila">http://www.fruitfly.org/TheBDGP/Drosophila</a>		
melanogaster BAC library was prepared by Kazuhiro Osoegawa and		
Aaron Mammoser in Pieter de Jong's laboratory in the Department of		
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,		
NY. The library is named RPCT-98 and was constructed by partial		
ECORI digestion of Drosophila DNA provided by the BDGP from the		
isogenic strain y2; cn bw sp, the same strain used for the BDGP's		
p1 and EST libraries. A more detailed description of the library		
and how to order individual BAC clones, the entire library, or		
filters for hybridization from the BACPAC Resource Center can be		
found at <a href="http://bdgpc.med.buffalo.edu/drosophila_bac.htm">http://bdgpc.med.buffalo.edu/drosophila_bac.htm</a> .		
FEATURES		
source	location/Qualifiers	
1..1001		
/organism="Drosophila melanogaster"		
/db_xref="taxon:7227		
/clone_lib="RPCT-98"		
/clone="BACR35406"		
/note="end : 77"		
BASE COUNT	266 a 209 c 148 g 202 t 176 others	
ORIGIN		
Query Match	1.7%: Score 37.4; DB 12; Length 1001;	
Best Local Similarity 39.8%: Pied. No. 11:		
Matches 72; Conservative 26; Mismatches 83; Indels 0; Gaps 0;		
Qy 1645 acctgtgcttggatcccgctgtgaagagatgcatgtcgcacgctgcatacgtg 1704		
Db 328 ACATATGACACCAACAGAGGCTCWMKAGAMAAATCTARRAAGCAACGARGGGM 387		
Qy 1705 gattctctggagcgagatctgtcttcgcgcagttgtttcatalaacgtttgttgtag 1764		
Db 388 GATRGTCGMRSTAAAAAKWTGCGCTKCGKATGTATTTGGAANAACMCTGTGGAAT 447		
Qy 1765 gggccagagaggtctactcaagaagaagaagttgctgtgtgagattgaagaatct 1824		
Db 448 AGCGCAACGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 507		
Qy 1825 t 1825		
Db 508 T 508		
RESULT 14		
LOCUS B87304		
DEFINITION RPCT11-28H13.TP RPCT-11 Homo sapiens genomic clone RPCT-11-28H13,		
DNA sequence.		
ACCESSION B87304		
VERSION B87304.1 GI:2928436		
KEYWORDS GSS.		
SOURCE human.		
ORGANISM Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
1 (bases 1 to 513)		
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,		
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and		

TITLE	Venter, J.C.
JOURNAL	Use of BAC End Sequences for Sequence-Ready Map Building (1998)
COMMENT	Unpublished (1998) Other GSSS: RPC11-28H13.TV Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Piter de Jong (pjter@jlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://Bacpac.med.buffalo.edu/ordering">http://Bacpac.med.buffalo.edu/ordering</a> ) or from Research Genetics ( <a href="http://www.tigr.org/tcdb/humgen/Bac_end_search.html">info@resgen.com</a> ). BAC end search page: <a href="http://www.tigr.org/tcdb/humgen/Bac_end_search.html">http://www.tigr.org/tcdb/humgen/Bac_end_search.html</a> Seq primer: SP6 Class: BNC ends
FEATURES	Location/Qualifiers
SOURCE	1..513 /organism="Homo sapiens" /db_xref="GDB:7510548" /db_xref="taxon:9606" /clone="RPC1-11-28H13" /clone_11b="RPC1-11" /sex="Male" /cell_type="Lymphocytes" /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPC11 Human Male BAC Library"
BASE COUNT	178 a 105 c 74 g 156 t
ORIGIN	
Query Match	1.7%; Score 37.2; DB 12; Length 513;
Best Local Similarity	56.6%; Pred. No. 8.1;
Matches	69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Dy	1883 catccatagacacatagtcttactcttggtggttcaccacctagatcatagacctt 1942
Dx	264 CAACTAAGAACAATTTCTCCAGTGTGGCGTCACAAATTCAACTCTCCTAGAAAT 323
Oy	1943 ctcaatctatttggcgcttgtttgcagatcgaigtaacacgcagcagaagcttggact 2002
Dz	324 GCCCTCTGCTAAAGGCTTGCTTAATCAATTAATTAACAACCACCTCATGAGCAGTTAACT 383
Oz	2003 gcc 2004
Dd	384 TC 385
RESULT 15	
CNS03F69/c	1067 bp DNA Linear GSS 17-MAY-2000
LOCUS	Tetradon nigroviridis genome survey sequence T7 end of clone
DEFINITION	02D1D7 of library G From Tetradon nigroviridis, genomic survey
ACCESSION	CNS03F69
VERSION	AL241290.1 GI:7962059
KEYWORDS	GSS: genome survey sequence.
SOURCE	Tetradon nigroviridis.
ORGANISM	Tetradon nigroviridis. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acantthomorpha; Acanthopterygii; Percormorpha; Tetraodoniformes; Tetraodontidae; Tetraodon. 1 (bases 1 to 1067) Roeser-Crollius,H.; Tallon,O.; Dasilva,C.; Fizames,C.; Fisher,C.; Bonneau,L.; Billault,A.; Queffelec,F.; Saurin,M.; Bernot,A. and Weissenbach,J.
REFERENCE	Characterization and repeat analysis of the compact genome of the
AUTHORS	freshwater pufferfish Tetradon nigroviridis
TITLE	Unpublished
JOURNAL	

```

REFERENCE      2 (bases 1 to 1067)
AUTHORS        Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
                Bernot,A., Fizames,C., Wincker,P., Brotier,P., Quetier,F.,
                Saulin,W. and Weissenbach,J.
TITLE          Human gene number estimate provided by genome wide analysis using
                Tetradon nigroviridis DNA sequence
JOURNAL        Unpublished
REFERENCE      3 (bases 1 to 1067)
AUTHORS        Genoscope.
TITLE          Direct Submission
COMMENT        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
                This sequence is a single read and was generated as part of a large
                scale clone-end sequencing project of the Tetradon nigroviridis
                genome. For more information, please take a look at
                http://www.genoscope.cns.fr/Tetradon.
FEATURES       Location/Qualifiers
                source          1..1067
                                /organism="Tetradon nigroviridis"
                                /db_xref="taxon:99883"
                                /clone="021D17"
                                /clone_1b="G"
                                /note="Genoscope sequence ID : CDBG021CB09LPI-end : T7"
BASE COUNT     356 a      185 c      268 g      257 t      1 others
ORIGIN
Query Match    1.7%: Score 37.2; DB 12; Length 1067;
Best Local Similarity 54.3%: Pred.No.14;
Matches 75; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 1702 tgggatttcttgagcgagatctgctccgagttgttcataacgtttggtcgt 1761
    || || || || || || || || || || || || || || || || || || ||
Db 624 TGATTTTATGTGACTGCAGTTGTGCTCTCCGCCGCTTTTAAATTTGATTGGT 565
QY 1762 agggggccctaacgagctactcaagcaatgaagaagtgctggtgtgattgacagaa 1821
    || || || || || || || || || || || || || || || || || || ||
Db 564 GGGGCTCACTGACTTTGTGTCAAGCAATGATACATGTTTGTGTGTCATATATTAGAA 505
QY 1822 tctttgagagattgctt 1839
    || || || || || || || || || || || || || || || || || || ||
Db 504 TACTTTGTGTAAAGTTT 487

```

Search completed: September 15, 2002, 22:53:08  
 Job time: 5077 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2002, 23:46:11 : Search time 58.43 Seconds  
(without alignments)  
629,222 Million cell updates/sec

Title: US-09-828-313-35

Perfect score: 1734  
Sequence: 1 MGITPSCVTVQGVYRVYR.....SAASQVQRNRGNKSLYMA 331

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	125.5	7.2	1381	22	ABR70001
2	116	6.7	912	16	ABR83442
3	105.5	6.1	2001	22	ABR20062
4	105.5	6.1	2001	22	ABR20063
5	105	6.1	560	22	AAU22961
6	104.5	6.0	377	22	AAE03643
7	104	6.0	1284	22	ABR58814
8	103.5	6.0	750	12	AAE15160
9	103.5	6.0	1321	17	AAE95994
10	103	5.9	2742	21	AAE23012
11	103	5.9	2783	13	AAE23962

12	103	5.9	2783	13	AAE23963	AFP-1 (Ala 2460 Val
13	103	5.9	2783	22	ABR82946	Human androgen rec
14	103	5.9	2842	15	ABR63508	Adenomatous polypo
15	103	5.9	2843	13	ABR26052	APC gene product 1
16	103	5.9	2843	15	AAE58634	Adenomatous polypo
17	103	5.9	2843	16	AAW11922	Adenomatous polypo
18	103	5.9	2843	18	AAW35392	Human adenomatous
19	103	5.9	2843	19	AAW76140	Human APC protein
20	103	5.9	2843	19	AAW76144	Human APC protein
21	103	5.9	2843	19	AAW38370	Human adenomatous
22	103	5.9	2843	21	AAE23011	Human APC protein
23	103	5.9	2860	15	AAE65507	Adenomatous polypo
24	103	5.9	2973	19	AAW76821	Human APC protein
25	103	5.9	2973	21	AAW70304	Human APC protein
26	103	5.9	2973	22	AAE72782	Protein used in ca
27	102	5.9	1231	22	ABR59466	Transcriptional ac
28	101.5	5.9	590	22	ABR58808	Drosophila melanog
29	101.5	5.9	795	22	ABR68471	Drosophila melanog
30	101	5.8	519	21	AAV44247	Human cell signal
31	101	5.8	679	19	AAW81739	M. tuberculosis im
32	101	5.8	679	19	AAW64372	Mycobacterium tube
33	101	5.8	679	20	AAV39169	M. tuberculosis an
34	101	5.8	679	20	AAV39026	Drosophila melanog
35	101	5.8	1540	22	ABR61913	Mouse YMK1 partial
36	100	5.8	382	21	AAV57448	Drosophila melanog
37	100	5.8	611	22	ABR68041	Arabidopsis thalia
38	100	5.8	632	21	AAE28283	Arabidopsis thalia
39	100	5.8	734	21	AAE28282	Arabidopsis thalia
40	100	5.8	736	21	AAE28281	Arabidopsis thalia
41	100	5.8	1130	20	AAV06526	Human WMT1, ortho
42	100	5.8	1222	21	AAE36453	Arabidopsis thalia
43	100	5.8	1257	21	AAE36452	Arabidopsis thalia
44	100	5.8	1275	21	AAE36451	Arabidopsis thalia
45	99.5	5.7	285	22	AAE94278	Human protein sequ

## ALIGNMENTS

RESULT 1	ABR70001 standard; Protein: 1381 AA.
XX	ABR70001:
AC	26-MAR-2002 (first entry)
XX	Drosophila melanogaster polypeptide SEQ ID NO 36795.
XX	pharmaceutical.
XX	Drosophila melanogaster.
XX	OS
XX	WO200171042-A2.
XX	27-SEP-2001.
XX	23-MAR-2001; 2001WO-US09231.
XX	23-MAR-2000; 2000US-191637P.
XX	11-JUL-2000; 2000US-0614150.
XX	(PEKE ) PE CORP NY.
XX	Venter JC, Adams M, Li PWD, Myers EW;
XX	WPI: 2001-656860/75.
XX	N-PSDB; ABL14104.
XX	New isolated nucleic acid detection reagent for detecting 1000 or more
XX	genes from Drosophila and for elucidating cell signalling and cell-cell
XX	interactions -
XX	interactions -

XX Disclosure: SEQ ID NO 36795; 21pp + Sequence Listing; English.

PS The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signaling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pcr\_sequences.

XX

SQ Sequence 1381 AA:

Query Match 7.2%; Score 125.5; DB 22; Length 1381;

Best Local Similarity 22.4%; Pred. No. 0.011;

Matches 72; Conservative 46; Mismatches 135; Indels 69; Gaps 15;

QY 30 SCGVHDLGHP-DYVCGSTPYTITNMAAEVL-----EYGVTFVCAPNAPFLER 83

DB 367 stlhdldlegeykfrkaenpyglsepsgesellfpdkrgltkpkasatrlagdekdk 426

QY 84 QPKVYHRSKILPRFSKGVHRELSPTHGSQSKRYFDYHVTMOQLSIRNGEPFH 143

DB 427 -pktagagmyvpr-----tktlsprrpqdaast-----gmsspkgspsaktr-pkpg 471

QY 144 LAGDRPSKH-LKLVIRHCLRALRLPRISIDMSPPNLSGALSPTATAKDETQWIL 202

DB 472 lldneqlthemsytschahk-----mdyrkpsalnadaaanpkltdssnpklnltl 523

QY 203 KSAARSELGMYTSKROEFLRRARRRKFAMKPVLOSISEMKVMEFHTPMAYRDSGSP- 261

DB 524 vtllapltdkavp-----spkaptptpslplklfnp-----kpsgapk 561

QY 262 -----PKNASTPRLSGPKNIS--PROVSVPORSSPPKVVSP-PROPAFVA----R 306

DB 562 dsrpgpkrpqlptprrmetpdkasppkrksisppnkqppplkspctpepikvtpallr 621

QY 307 TASKIS-AA5QOVORNRGNKS 327

DB 622 saepvqlgvnqvrrfsgqls 643

RESULT 2

AA85442

ID AAR85442 standard; Protein; 912 AA.

XX

AC AAR85442;

XX

DT 19-FEB-1996 (first entry)

XX

DE Bovine brevican core protein.

XX

KW Brevican; chondroitin sulphate proteoglycan; glial cell; axon;

KW neurofibromatosis; gliosis.

XX

OS Bos taurus.

XX

FH Key Location/Qualifiers

FT Peptide 1..34

FT /label= Sig\_peptide

FT Region 35..158

FT /label= Immunoglobulin-like\_loop

FT Modified-site 130

FT /label= N-glycosylation\_site

FT Region 159..353

FT /label= Protein-like\_tandem\_repeats

FT Modified-site 337

FT /label= N-glycosylation\_site

FT

FT Region 354..648

FT /label= Central\_region

FT Region 649..684

FT /label= EGF-like\_repeat

FT Region 685..814

FT /label= Lectin-like\_region

FT Region 815..875

FT /label= CRP-like\_region

PN W09526201-A1.

XX

PD 05-OCT-1995.

XX

PD 27-MAR-1995; 95WO-US03747.

PF

XX

PR 28-MAR-1994; 94US-0219642.

XX

PA (JOL-) LA JOLIA CANCER RES FOUND.

XX

PI Shimonaka M, Watanabe K, Yamada H, Yamaguchi Y;

PI WPI: 1995-351200/45.

DR N-PSDB; AAT05627.

XX

PT Mammalian brevican protein - directs/inhibits axonal growth, used in

PT treatment of neuro-fibromatosis and in detection of gliosis

PS Claim 2; Page 42-44; 73pp; English.

CC 2 overlapping cDNA clones obtd. from a bovine brain cDNA library

CC were combined (see AAT05627) to provide cDNA coding for bovine

CC brevican (AAR85442). The brevican can be used to modulate axonal

CC growth and to raise antibodies used in detection of gliosis.

XX

SQ Sequence 912 AA:

Query Match 6.7%; Score 116; DB 16; Length 912;

Best Local Similarity 24.2%; Pred. No. 0.051;

Matches 54; Conservative 26; Mismatches 101; Indels 42; Gaps 8;

QY 140 PEPILADRSKILKLVIRHCLRALRLPR-----ISIDME-----SPLPNIS 183

DB 366 pashlaadale---alvtvletleelkipeaveesrgalytpledggsgspdepd 424

QY 184 GEALSPATAKDETQWILKSAARSELGMYTSKROEFLRRARRRRK-----FA 232

DB 425 aea--prtllefe-tgslvprpysseegkvleqeekeeeeeevedeaalwa 481

QY 233 WKPVLOSISEMKVMEFHTPMAYRDSGSPPKNA-----STPSLGPKNISPPROVSVPO 286

DB 482 wpselsldpeapldptepypesltqasppyratlgpyvspppydeaprrprvlgppt 541

QY 287 RSSPPPKN--VSPPPQPAFVAKTASKYSAA5QOVORNRGNK 326

DB 542 kltpcpregnlaspptslvgareleeltgslgapsaprgese 584

RESULT 3

AAB20062

ID AAB20062 standard; Protein; 2001 AA.

XX

AC AAB20062;

XX

DT 23-APR-2001 (first entry)

XX

DE Arabidopsis thaliana silencing gene-encoded protein.

XX

KW Gene silencing; silencing gene; MOM.

XX

OS Arabidopsis thaliana.

XX

FH Key Location/Qualifiers

FT	Region	177..350	/note="Internal repeat region"
FT	Peptide	362..367	/note="nuclear localization motif"
FT	Binding-site	460..467	/note="ATP/GTP-binding motif"
FT	Region	479..719	/note="region of homology to ATPase/helicase family SWI2/SNF proteins"
FT	Misc-difference	705	/note="Iys in ecotype Columbia"
FT	Peptide	832..838	/note="nuclear localization motif"
FT	Peptide	838..862	/note="nuclear localization motif"
FT	Domain	995..1015	/note="predicted membrane-spanning domain"
FT	Misc-difference	1219	/note="Asp in ecotype Columbia"
FT	Region	1462..1672	/note="Internal repeat region"
FT	Region	1848..1894	/note="Internal repeat region"
FT	Domain	1899..1941	/note="actin-binding domain"
XX			
PN		MO200100801-A2.	
PD	04-JAN-2001.		
PE	21-JUN-2000; 2000WC-EP05761.		
XX			
PR	23-JUN-1999; 99GB-0014623.		
XX			
PA	(NOVS ) NOVARTIS AG.		
PA	(NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.		
XX			
P1	Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;		
XX			
DR	NP1; 2001-137952/14.		
DR	N-PSDB; AAA89353, AAA89354.		
XX			
PT	Novel gene encoding a protein that controls gene silencing, in		
PRT	particular silencing of plant genes -		
XX			
P5	Claim 1; Page 32-39; 48pp; English.		
CC			
CC	The present sequence is that of Arabidopsis thaliana protein MOM,		
CC	which controls gene silencing, particularly gene silencing. It		
CC	is encoded by a novel silencer gene (see AAA89353) of A. thaliana		
CC	ecotype zurich. The protein exhibits partial similarity with		
CC	ATPase/helicase proteins of the SWI2/SNF2 family. The sequence		
CC	(see AAA20063) encoded by the silencer gene of wild-type A.		
CC	thaliana ecotype Columbia. Gene silencing is useful as a molecular		
CC	tool for regulating gene expression.		
XX			
SQ	Sequence 2001 AA;		

```

Query Match Similarity 6.1% Score 105.5: DB 22: Length 2001:
Best Local Similarity 20.4%: Pred. No. 1.5:
Matches 82: Conservative 56: Mismatches 128: Indels 135: Gaps 19:

QY 2 GLLPESCTVQGYVYVYVPPGHVENSKSCSHDILLGNPDYVYCGSTPTTTNNMAAE 61
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1645 gnhpdcvaileldlctsvaephsg--sdcc--emelesepqyvetstaaflfiegveh 1700
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 62 VLEGVYTVFCATPNAOPEFLEROKPVVHFGSKIL-----PRFSKHGVHRELRSK 111
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1701 --sagvtalv-----psllngtletqavcpvqldpfvfdnprflth--elekl 1743
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 112 THGQGSRRV--DYHSTVYMOLESINNEBPENHLAGRPSKHKLVEIR--HCLTA--L 165

```

```

Db      1744 rreesnksfleeekssllkæle-----rkmaeqgefefirkfheesehnt 1769

OY      166 RLRPRSID-----LMSEPLPNISGALSPTATKKEIDITOMLKRAARSELGVYVK 216
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1790 rftklexkdmlytmklllaafstckcdkxwspgsaprgyrlqlqlqraaysalrnylap 1849
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      217 ROEFLYLRARRRRRRFAMKPVLOSIEMKPFMEFHTPMAYRDSGSP-----P 262
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1850 qg-----lqgsfipapal_vsaaplqgsfipadpapgaplqpgasfp 1890
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      263 KNASTPS-----LPQPK-----NISPPOVS-----VQORSPPKPNVS 297
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1891 ssvstspallllfawcempgprqltsmlnaprpsvtpatnpjrlapaphlnsyrpsstp 1950
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      298 -----PPQPAFVARTSKYSKSAASOOVRRNRGNKAKSL 328
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1951 valatpssvppq-----alysasvsiqqqdeqpgqsl 1984

```

RESULT	4	
ABB20063		
ID	ABB20063	standard; Protein; 2001 AA.
XX	ABB20063;	
AC		
XX		
DT	23-APR-2001	(first entry)
XX		
DE	Arabidopsis thaliana	silencing gene-encoded protein.
XX		
KW	Gene silencing; silencing gene; MOM.	
XX		
OS	Arabidopsis thaliana.	
XX		
FH	Key	Location/Qualifiers
FT	Region	177..350
FT		/note="Internal repeat region"
FT	Peptide	362..367
FT		/note="nuclear localization motif"
FT	Binding-site	460..467
FT		/note="ATP/GTP-binding motif"
FT	Region	479..719
FT		/note="region of homology to Arpase/helicase family SWI2/SNF proteins"
FT		
FT	Misc-difference	705
FT		/note="Met in ecotype Zurich"
FT	Peptide	832..838
FT		/note="nuclear localization motif"
FT	Peptide	858..862
FT		/note="nuclear localization motif"
FT	Domain	995..1015
FT		/note="Predicted membrane-spanning domain"
FT	Misc-difference	1219
FT		/note="Glu in ecotype Zurich"
FT	Region	1462..1672
FT		/note="Internal repeat region"
FT	Region	1848..1894
FT		/note="Internal repeat region"
FT	Domain	1899..1941
FT		/note="actin-binding domain"
XX		
PN	W0200100801-A2.	
XX		
PD	04-JAN-2001.	
XX		
XX	21-JUN-2000; 2000MO-EP05761.	
PF		
XX		
XX	23-JUN-1999; 99GB-0014623.	
XX		
PA	(NOVS ) NOVARIS AG.	
PA	(NOVS ) NOVARIS-ERFINDUNGEN VERW GES MBH.	
PI	Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;	

XX WPI: 2001-137952/14.  
DR N-PSDB; AAA89385.  
XX  
XX Novel gene encoding a protein that controls gene silencing, in  
XX particular silencing of plant genes -  
XX  
XX Claim 8; Page -: 48pp; English.  
XX  
XX The present sequence is that of Arabidopsis thaliana protein MOM,  
CC which controls gene silencing, particularly gene silencing. It  
CC is encoded by a novel silencer gene (see AAA89385) of A. thaliana  
CC ecotype Columbia. The protein exhibits partial similarity with  
CC ATPase/helicase proteins of the SWI2/SNF2 family. The sequence  
CC differs at amino acids 705 and 1219 compared with the protein  
CC (see AAA20062) encoded by the silencer gene of Arabidopsis thaliana  
CC ecotype Zurich. Gene silencing is useful as a molecular tool for  
CC regulating gene expression.  
CC Note: the present sequence is not shown in the specification but  
CC is derived from the A. thaliana mutant silencer gene protein  
CC sequence given in AAB20062.  
XX  
XX  
SQ Sequence 2001 AA:  
  
Query Match 6.1%; Score 105.5; DB 22; Length 2001;  
Best Local Similarity 20.4%; Pred. No. 1.5;  
Matches 82; Conservative 56; Mismatches 128; Indels 135; Gaps 19;  
  
QY 2 GLTFSCVTVQGVYKRVYVDGHEVNLKSCSVHDLGNDYVYCGSPYRTITNMAAE 61  
DB 1645 gnhpdtavnlvgidntsvaephlsq-sdac---emealsepgqvverstaniifhegyeh 1700  
QY 62 VLEGVTVFVCATPNAPLELRQPKVHRGSKIL-----PRFSKHGVHRELKSP 111  
DB 1701 --sagvcalv-----pallngteqlavgpvpqipfvyfnpflh--elekl 1743  
QY 112 THGSOOSRKVF-DYHSVTMOQLSEIRNEGPEPHLAGRPSEKHLKLVFR--HCLRA--L 165  
DB 1744 rresenakkkfiekkskilaee-----rkmaevgaefrrkkfhevaeht 1789  
QY 166 RLPRITSD-----LMESPLPMLSGEALSPYNAKKDITOMILKSAASELGMYSK 216  
DB 1790 rtkiekdkhvlmnhkllanaflskctdkkvspgprgkldqiaqraqvaaalnnylap 1849  
QY 217 ROEFFYLRARRRRRKFAMKPVLOSISEMKPVMEFRTPMAYRDSGSP-----P 262  
DB 1850 qg-----lqassfpapal-vsapldlqssffpdpapdlqpgassfp 1890  
QY 263 KMASTPS-----LFGPK-----NISPPROVS-----VPQRSSPPPKNVSP 297  
DB 1891 ssvavrpalllnfavcpmpqprqplslmlaptpatnpglrsphnlinsyrpsastp 1950  
QY 298 -----PPQPAFVARTASKYGAASQGVORNRGNKSL 328  
DB 1951 valatpctssvppq-----altysavslqgqgqgqgqsl 1984  
  
RESULT 5  
AAU02961 standard; Protein: 560 AA.  
ID AAU02961;  
XX  
AC AAU02961;  
XX  
XX 18-DEC-2001 (first entry)  
XX  
DE Novel human enzyme polypeptide #47.  
XX  
XX Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;

KW nephrotropic; anticoagulant.  
XX  
XX Homo sapiens.  
XX  
XX WO200155301-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001WO-0501239.  
XX  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0227182.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227909.  
XX 30-AUG-2000; 2000US-0228287.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 05-SEP-2000; 2000US-0230457.  
XX 06-SEP-2000; 2000US-0230458.  
XX 06-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.  
XX 14-SEP-2000; 2000US-0232400.  
XX 14-SEP-2000; 2000US-0232401.  
XX 14-SEP-2000; 2000US-0233063.  
XX 14-SEP-2000; 2000US-0233064.  
XX 14-SEP-2000; 2000US-0233065.  
XX 14-SEP-2000; 2000US-0234223.  
XX 21-SEP-2000; 2000US-0234223.  
XX 21-SEP-2000; 2000US-0234274.  
XX 25-SEP-2000; 2000US-0234997.



Human: extracellular matrix and cell adhesion molecule; XMAD;  
 gene therapy; genetic disorder; adrenoleukodystrophy; leukemia;  
 Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;  
 sickle cell anaemia; thalassemia; autoimmune disorder; adenocarcinoma;  
 inflammatory disorder; acquired immune deficiency syndrome; AIDS;  
 Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;  
 Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;  
 glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;  
 osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;  
 infection; cell proliferative disorder; actinic keratosis; myeloma;  
 atherosclerosis; neotropic; anticonvulsant; antithyroid; nephrotropic;  
 neuroprotective; dermatological.

Homo sapiens.  
 Key Location/Qualifiers  
 Region 20..211  
 /note="Leucine rich repeat"

MO200142285-A2.

14-JUN-2001.

05-DEC-2000: 2000MO-US32990.

10-DEC-1999: 990S-0172852.

16-DEC-1999: 990S-0172354.

(INCY-) INCYTE GENOMICS INC.

Yue H, Tang YT, Lal P, Burford N, Azimzal Y, Patterson C;  
 Baughn MR, Lu DM, Shah P, Au-Young J;

WPI: 2001-381632/40.

N-PSDB: AAD08051.

New human extracellular matrix and cell adhesion molecules and  
 polynucleotide sequences encoding them, useful for diagnosis,  
 prevention, treatment of genetic, autoimmune and cell proliferative  
 disorders -

Claim 1; Page 100-101; 135pp; English.

The present sequence is a human extracellular matrix and cell  
 adhesion molecule (XMAD). The XMAD is used for screening a compound for  
 effectiveness as an agonist or antagonist of XMAD. The identified agonist  
 or antagonist are used for treating a disease or condition associated  
 with decreased or increased expression of functional XMAD. The  
 polynucleotides encoding XMAD are useful in somatic or germline gene  
 therapy to correct a genetic deficiency, to express a conditionally  
 lethal gene product and to express a protein which affords protection  
 against intracellular parasites and also for diagnosis of disorders  
 associated with expression of XMAD. They are also used for generating  
 hybridisation probes useful in mapping the naturally occurring genomic  
 sequences and to create knock in humanised animals (pigs) or transgenic  
 animals (mice or rats) to model human diseases. Oligonucleotide or longer  
 fragments derived from the polynucleotide sequences may be used as  
 elements on a microarray. Antibodies which specifically bind XMAD may be  
 used for the diagnosis of disorders associated with the expression of  
 XMAD, or in assays to monitor patients being treated with XMAD. Diseases  
 diagnosed, prevented or treated include genetic disorders such as  
 adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's  
 disease, myotonic dystrophy, sickle cell anaemia, thalassemia,  
 autoimmune/inflammatory disorders such as acquired immune deficiency  
 syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,  
 atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,  
 glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,  
 osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,  
 bacterial, fungal, parasitic, protozoal and helminthic infections and  
 cell proliferative disorders such as actinic keratosis, arteriosclerosis  
 and cancer including breast, bladder, bone marrow, brain and uterus  
 cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.

SO Sequence 377 AA;

Query Match 6.0%; Score 104.5; DB 22; Length 377;  
 Best Local Similarity 19.3%; Pred. No. 0.19;  
 Matches 52; Conservative 31; Mismatches 101; Indels 85; Gaps 9;

98 ESKGCVHRELRSTHSGSQSRKVPD---YHSVTMOOLESTRN-----137

19 fsegredhelrthrhghspkshlelphygllahqpfifrsycklrcpklppsp 78

138 ----EGPEPHLGDPRPSKHLKLVFIRHCLRLRLPRISIDLMESPUNSGEALSPRTA 193

79 nkpkfppph-----qpkh-----pkksvvpplva 107

194 KDEITOMLKSAAS-----ELGMYSKROEYLRARRRRKFAWKPVLOSISE 242

108 ttqipsvclfpasltkltpnvltfpnatlsrenvntassvatlapynpdpqdtca 167

243 MKPVMFHTPMAYRDSGSPPKNASRPSLPG-----PKNISPPHOVS-----VPQ 286

168 appipsattp-appssappettaappptattgappssappettaappptattppp 226

287 RSPPKNVSPPPQPAFVARTASKYSAAS 315

227 sssappettaappcps--attcpajsssa 253

RESULT 7

ABBS8814

ID ABBS8814 standard; Protein; 1284 AA.

XX ABBS8814;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 3234.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmacological.

XX Drosophila melanogaster.

XX WO20011042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX N-PSDB: ABL02917.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure: SEQ ID NO 3234; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX (ABBS7737-ABBS72022).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 1284 AA:

Query Match 6.0%; Score 104; DB 22; Length 1284;  
Best Local Similarity 20.6%; Pred. No. 1.1;  
Matches 79; Conservative 43; Mismatches 107; Indels 154; Gaps 19;

QY 16 RVVYPDGHEVNLKSCSVH-----DLLGNPDYV-----CGSTPYTTNNM 57  
DB 631 rmlfrngqakhrstakgmnlvntwgrllkrrapnvmgsaagsalqgspmvvgasr 690  
QY 58 AAEEVLEGVTVYVCATPNAQPFLEKOPKVVHRGSKILPRFSKHGVHRELRSPTHSQO 117  
DB 691 deesp-----lsrtpsdaipepy-----tpgega 717  
QY 118 SRKVDYHSVTMOLESIRNEGPEP---HLGDRP-SKHKLIVIRHCLRALPLRISTD 173  
DB 718 qpletdpdaglnelvet-pgeypdaasglsgmrplsmhg-----gisvl 762  
QY 174 LMESE-----LPNLS-----GEALSPRA-----TAKDEITOMILKS 204  
DB 763 ppsppvatgaagrpnlslqmpgaskgvlqgrtaaptlappppvlkatastplldq 822  
QY 205 AARSELG-MYVSKROEFLRRARRRRKFAWKPVLOSISEMKPVMEFTTPMAVYRDSGSPK 263  
DB 823 evlppqlglllygssgqgqyqgs-----spilq-----epatpcliynsaa--- 862  
QY 264 NASTSTLPGKPNISPPROSVFQRSSPPPKNVSPPPOPAF-----VARTASKYSA 313  
DB 863 -agapqfifqp-----aqrqekqp---pqgqpiyangyelnvakaaasvysp 906  
QY 314 AS-----OOVORNRGNNAKSL 328  
DB 907 sssttsnsngqgqgrnnrvargl 929

RESULT 8

ID AAR15160 standard; Protein; 750 AA.

XX AAR15160;  
AC AAR15160;  
DT 17-DEC-2001 (updated)  
DT 28-FEB-1992 (first entry)  
XX FB29 chicken c-ski protein.  
DE  
XX Transgenic; muscle growth; muscle degeneration; obesity.  
XX Gallus domesticus.  
OS  
XX USN7620415-N.  
XX  
XX PD 05-NOV-1991.  
XX  
XX PF 03-DEC-1990; 900S-0620415.  
XX  
XX PR 03-DEC-1990; 900S-0620415.  
XX  
XX PA (USSH ) NAT INST OF HEALTH.  
XX  
XX PI Hughes S, Suttrave P;  
XX  
XX DR WPI; 1991-361684/49.  
XX  
XX DR N-PSDB; AAQ14799.  
XX  
XX PT DNA encoding chicken C-ski proteins - used to produce transgenic  
XX PT animals having increased muscle growth, and to treat muscle  
XX PT degeneration or obesity

XX Disclosure; Fig 2; 55pp; English.  
XX

CC This is the amino acid sequence of the chicken c-ski protein FB29.  
CC It may be produced in transgenic animals, using a DNA construct, to  
CC stimulate muscle growth or prevent muscle degeneration. It can be  
CC used to produce domestic livestock with increased muscle size and  
CC decreased fat tissue. It may also be used to treat serious muscle  
CC injury or muscle degenerative diseases, or patients suffering from  
CC obesity.

CC (Note: Revised entry submitted to correct the patent number format of  
CC US Government-owned NTIS applications to prevent clashes with ongoing US  
CC granted patent numbers. For further information please visit the Derwent  
CC web site at [www.derwent.com/dwpi/updates/ntis\\_us.html](http://www.derwent.com/dwpi/updates/ntis_us.html).)  
XX

SO Sequence 750 AA:

Query Match 6.0%; Score 103.5; DB 12; Length 750;  
Best Local Similarity 21.0%; Pred. No. 0.6;  
Matches 81; Conservative 40; Mismatches 134; Indels 131; Gaps 17;

QY 1 MGLTFSCVTVGTVRVVYPOGHEVNLKSCSVHDDLGNPDYVCGSTPYTTNNMAAE 60  
DB 147 mgltfscv-----cyltktlaerlc-- 169  
QY 61 EVLEGVTVYVCATPNAQ-----PFLEKOPKVVHRGSKILPRFSK-HGVHRELRS 110  
DB 170 nallygqly-----phckkqfjastleleltekstkyhes-----fgkckgllvlpelys 219  
QY 111 -PTHSQO--SRKVDYHSVTMOLESIRNE-----GPE--- 141  
DB 220 nptaacqgclhctlmypbhkfvvshkslqstcnhwgfdsanvtylllsqdytqkeeka 279  
QY 142 --PILGADRPSE-HLKLIVIRHCLRALPLRISI-----DLMESEPLNLT 182  
DB 280 tlpplldemkekfdynnkkykkrpnrspvqlrpnkmfklmldpaagsavylqpdq 339  
QY 183 SGEALSPRTAKDEITOMILKSARSELGMVSKROEFLRRAR-----RRK 230  
DB 340 neypsdpascktkldsdapsqspaste-----kekgswlrlsssnkslqcvhprql 394  
QY 231 FAKKPVLOSISEMKPVMEFTTPMAVYRDSG-----SPPKNASTP--SLPSP--KNISPPRO 281  
DB 395 salppwspavsanekelstlhpallrdsfysyksenavapvalappaqkvvsnpnc 454  
QY 282 VSVFQRSSPPPKNVSPPPOPAFVART 307  
DB 455 atvpspssepssaaqppkrrhaet 480

RESULT 9

ID AAR96994 standard; Protein; 1321 AA.

XX AAR96994;  
AC AAR96994;  
DT 16-AUG-1996 (first entry)  
XX  
XX DE Mouse IRS-2.  
XX  
XX KM IRS-2; insulin receptor substrate-2; diabetes; therapy;  
XX KM diagnosis; vector; antibody; transgenic animal.  
XX  
XX OS Mus sp.  
XX  
XX Key Location/Qualifiers  
XX 30..141  
XX Domain  
XX  
XX FT /label= IH1  
XX FT /note= "IRS-homology domain I"  
XX FT 190..366  
XX FT /label= IH2  
XX FT /note= "IRS-homology domain II"





XX (ICIL ) IMPERIAL CHEM IND PLC.  
 PA (UTIO ) UNIV JOHNS HOPKINS.  
 PA (UTAH ) UNIV UTAH.  
 PA (CANC-) CANCER INST.

XX Carlson M, Groden J, Joslyn G, Kinzler K, Markham AF, Anand R;  
 PI Albertsen H, White RL, Thliveris A, Nakamura Y, Vogelstein B;  
 PI Hedge PJ;

XX WPI: 2000-565003/52.  
 DR N-PSDB; AAA93450.

XX Detecting Adenomatous Polyposis Coli (APC) protein in a sample for  
 PT diagnosing cancers, involves contacting the sample with antibodies that  
 PT specifically bind to APC protein and detecting the complex formed -  
 XX Example 7; Fig 7A1-7W; 125pp; English.

XX The invention relates to a novel method for detecting Adenomatous  
 CC Polyposis Coli (APC) protein in a sample. The method involves  
 CC contacting the sample with antibodies which specifically binds to the  
 CC 2843 amino acid form of the human APC protein, or to a mutant APC  
 CC protein, and detecting an APC-antibody complex. Mutations in the APC  
 CC gene play a role in tumorigenesis, indicating that it is a tumour  
 CC suppressor gene. It is located on chromosome 5q21, which corresponds to  
 CC the FAP (familial adenomatous polyposis) locus. FAP is an autosomal  
 CC dominant inherited disease in which affected individuals develop  
 CC hundreds to thousands of adenomatous polyps in the colon and rectum,  
 CC some of which progress to malignancy. The FAP locus is often found to  
 CC be deleted in sporadic (i.e., non-familial) adenomas and carcinomas, and  
 CC chromosome 5q deletions have also been observed in tumours of the lung,  
 CC breast, colon, rectum, bladder, liver, sarcomas, stomach, and prostate,  
 CC and in leukaemias and lymphomas. Although the FAP locus contains  
 CC several other genes such as RER, TBL, TB2, and MCC, it is thought that  
 CC mutations in the APC gene play a key role in the development of FAP and  
 CC sporadic tumours. The method is useful for detecting APC protein and its  
 CC serum or a tumour sample. The method is useful for diagnosing or  
 CC prognosing neoplastic tissue, for detecting a genetic predisposition to  
 CC cancer, for detecting germline and somatic alteration of wild-type APC  
 CC genes, and for testing therapeutic agents for the ability to suppress  
 CC tumours. The present sequence represents a 2742 amino acid splice  
 CC variant of the human APC protein. This variant is less abundant than the  
 CC 2843 amino acid variant (AAB23011).

XX Sequence 2742 AA;

Query Match 5.9%; Score 103; DB 21; Length 2742;

Best Local Similarity 22.3%; Pred. No. 4; Mismatches 107; Indels 46; Gaps 11;

Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PPLEKQKRVNHRGSKILPRFSKHGVNRELSPHSGQSKRYFDYHSVTWOOLSEISINE 138  
 DB 2051 pdgeekprtsmkprrllkpyekstletkkiesekgikgkv--yksiiltgkvs--- 2104

QY 139 GPEPHLAGDRPSKHLVFIHQLRALRIPRISIDLESPLNL--SGEALSPRTAKDE 196  
 DB 2105 --nseisgq-----mkqpqlga-nmpsisrgtcmihlpgvrnssstspvskkpp 2151

QY 197 ITQMLTKSARSELGMVSKROEFLRARRRRKFAKPVLOSLISEKRYMEFHRPMATR 256  
 DB 2153 lktpsakspsegqat--tspr-----gakkpsvk--selspvar-qtsqig 2193

QY 257 DSGSPPKNA--STPSLPGPKNISPPROVSVPORSSPPPKNVSPPOPAFVARTSKSYSA 313  
 DB 2194 sskapsrsgsdstpsrpaqpsrplsqpsgrnsgisppmklsqllprtspspta 2253

QY 314 ASQGVQRNRGMAKSLY 329

DB 2254 st-----kssgsgkmsy 2265

RESULT 11

AAR23962

ID AAR23962 standard; Protein; 2783 AA.

XX AAR23962;

XX 06-JAN-1993 (first entry)

XX AFP-1.

XX Enhancer; alpha fetoprotein; homeodomain; zinc-finger.

XX Homo sapiens.

XX Key

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

DNA encoding protein binding to alpha-fetoprotein gene enhancer -  
 useful for prodn. of biological active protein

PS Claim 1; Page 7; 24pp; English.  
 CC This protein specifically binds to the enhancer of the alpha-  
 CC fetoprotein gene. The DNA encoding this protein may be useful for  
 CC the prodn. of biologically active proteins by inserting the DNA into  
 CC an expression vector and co-transfecting animal cells with another  
 CC expression vector in which a gene for the biologically active protein  
 CC is placed under the control of alpha-fetoprotein gene enhancer and  
 CC promoter.  
 CC  
 XX Sequence 2783 AA:  
 SQ

Query Match 5.9%; Score 103; DB 13; Length 2783;  
 Best Local Similarity 28.0%; Pred. No. 4.1;  
 Matches 44; Conservative 20; Mismatches 49; Indels 44; Gaps 8;  
 QY 171 SIDME-----SPLPNSGEALSPTA--TAKDEITOMILKSAARSELGMVSKRQ 218  
 Db 1454 smdamelltpstsscsctmpsqaysapapsanntasafqlq---taaeelatfnskte 1510  
 QY 219 EFYLRRARRRRKFAW-----KPVLOSISEMKPVMEFHTPMAYRDSGSP 262  
 Db 1511 -----agdekplaieapsagpnqtgkqgqkpelq-qqeqpeqkntlpqg-----k1p 1558  
 QY 263 KNASTPSLPGPKNISPPROVSVPORSSPPPKNVSP 299  
 Db 1559 qlvslpslppppqppqcp1pq-sspspsqlshlp 1594

## RESULT 12

AA823963  
 ID AAR23963 standard; Protein: 2783 AA.

AC AAR23963;  
 DT 06-JAN-1993 (first entry)  
 XX  
 DE AFP-1 (Ala 2460 Val).

XX Enhancer: alpha fetoprotein; homeodomain; zinc-finger.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH Domain 1231..1291  
 FT /note= "homeodomain"  
 FT Domain 1328..1388  
 FT /label= homeodomain  
 FT Domain 1727..1786  
 FT /label= homeodomain  
 FT Domain 2032..2091  
 FT /label= homeodomain  
 FT Domain 67..99  
 FT /label= zinc-finger  
 FT Domain 123..155  
 FT /label= zinc-finger  
 FT Domain 171..203  
 FT /label= zinc-finger  
 FT Domain 306..366  
 FT /label= zinc-finger  
 FT Domain 443..476  
 FT /label= zinc-finger  
 FT Domain 484..543  
 FT /label= zinc-finger  
 FT Domain 628..660  
 FT /label= zinc-finger  
 FT Domain 679..711  
 FT /label= zinc-finger  
 FT Domain 1066..1097  
 FT /label= zinc-finger  
 FT Domain 1411..1442  
 FT /label= zinc-finger

FT Domain 1613..1643  
 FT /label= zinc-finger  
 FT Domain 1794..1825  
 FT /label= zinc-finger  
 FT Domain 2107..2139  
 FT /label= zinc-finger  
 FT Domain 2540..2571  
 FT /label= zinc-finger  
 FT Domain 2606..2638  
 FT /label= zinc-finger

PN EP487229-A.  
 XX  
 XX 27-MAY-1992.  
 PD  
 XX  
 XX 07-NOV-1991; 91EP-0310334.  
 PF  
 XX 07-NOV-1990; 90JP-0301412.  
 PR  
 XX (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PA  
 XX  
 XX Higashio K, Morinaga T, Tamaoki T, Yasuda N;  
 DR WPI: 1992-176828/22.  
 DR N-PSDB; AA024829.  
 XX  
 XX DNA encoding protein binding to alpha-fetoprotein gene enhancer -  
 PT useful for prodn. of biological active protein  
 XX  
 XX Claim 3-4; Page 7; 24pp; English.

CC This protein sequence is as the sequence given in AAR23962 except  
 CC that Ala at position 2460 is replaced by Val.  
 CC This protein specifically binds to the enhancer of the alpha-  
 CC fetoprotein gene. The DNA encoding this protein may be useful for  
 CC the prodn. of biologically active proteins by inserting the DNA into  
 CC an expression vector and co-transfecting animal cells with another  
 CC expression vector in which a gene for the biologically active protein  
 CC is placed under the control of alpha-fetoprotein gene enhancer and  
 CC promoter.  
 CC  
 XX Sequence 2783 AA:

Query Match 5.9%; Score 103; DB 13; Length 2783;  
 Best Local Similarity 28.0%; Pred. No. 4.1;  
 Matches 44; Conservative 20; Mismatches 49; Indels 44; Gaps 8;

QY 171 SIDME-----SPLPNSGEALSPTA--TAKDEITOMILKSAARSELGMVSKRQ 218  
 Db 1454 smdamelltpstsscsctmpsqaysapapsanntasafqlq---taaeelatfnskte 1510  
 QY 219 EFYLRRARRRRKFAW-----KPVLOSISEMKPVMEFHTPMAYRDSGSP 262  
 Db 1511 -----agdekplaieapsagpnqtgkqgqkpelq-qqeqpeqkntlpqg-----k1p 1558  
 QY 263 KNASTPSLPGPKNISPPROVSVPORSSPPPKNVSP 299  
 Db 1559 qlvslpslppppqppqcp1pq-sspspsqlshlp 1594

## RESULT 13

AA82946  
 ID AAB82946 standard; Protein: 2783 AA.

AC AAB82946;  
 XX  
 XX 21-DEC-2001 (first entry)  
 DT Human androgen receptor trapped protein (ART).  
 XX  
 DE Androgen receptor trapped protein; ART: human;  
 XX  
 KW transcriptional coregulatory protein; antagonist;

KM Screening: prostate cancer; benign prostatic hyperplasia;  
 XX hair loss; alopecia; therapy.

OS Homo sapiens.

PN WO200172332-A1.

PD 04-OCT-2001.

XX 26-MAR-2001; 2001WO-US09378.

XX 24-MAR-2000; 2000US-191768P.

XX 15-AUG-2000; 2000US-225618P.

XX (UYNV ) UNIV NEW YORK STATE.

PI Garabedian M, Taneja S, Hittelman A, Markus S;

DR MPI: 2001-611582/70.

DR N-PSDB; AAR63508.

PT New androgen receptor transcriptional coregulatory proteins that  
 PT interact with androgen receptor to regulate androgen-dependent gene  
 PT expression, useful for producing antibodies to treat androgen dependent  
 PT diseases

PS Claim 31; Page 99-107; 113pp; English.

CC The present sequence is that of a novel human androgen receptor  
 CC trapped protein (ART), which belongs to a new class of proteins  
 CC that interact with the N-terminus of the androgen receptor as  
 CC androgen receptor transcriptional coregulatory (i.e. coactivator)  
 CC proteins, regulating androgen-dependent gene expression. To identify  
 CC proteins that interact with the androgen receptor N-terminus,  
 CC a modified yeast two-hybrid system that allowed the identification  
 CC of factors expressed in the prostate and which associated with  
 CC transcriptional activators was used. An androgen-stimulated LNCaP  
 CC prostate cancer cell cDNA library fused to the LexA DNA binding  
 CC domain was screened for proteins that interacted with the androgen  
 CC receptor N-terminal transcriptional activation domain (amino acid  
 CC residues 18-500). From approximately 1 million library  
 CC transformants, 8 clones were isolated, and termed ARTs, for  
 CC androgen receptor trapped. ARTs are useful for producing  
 CC antibodies (or molecules having the binding portion of an antibody)  
 CC which can be administered to patients to treat androgen dependent  
 CC diseases, especially prostate cancer, benign prostatic hyperplasia  
 CC or androgen dependent hair loss. They can be used to identify  
 CC inhibitors disrupting interaction between an androgen receptor and  
 CC an androgen receptor transcriptional coregulatory protein, useful  
 CC to inhibit such interaction (all claimed) therapeutically.

XX Sequence 2783 AA;

Query Match 5.9%; Score 103; DB 22; Length 2783;

Best Local Similarity 28.0%; Pred. No. 4.1;

Matches 44; Conservative 20; Mismatches 49; Indels 44; Gaps 8;

OY 171 SIDLME-----SPJPNLSGEALSPTA--TAKDEITOMILKSAARSELGMVSKRO 218

DB 1454 smdameellftstssctpmqsqayapansantassafql---laeeelatinste 1510

OY 219 EFYLRARRRRRRKRAM-----KPYLOSISMKKVMFFHTPMAYRDSGSEP 262

DB 1511 -----agdekpklaeapsqpnqtekgqgkpelq--gqeqgqetnlpqg-----klp 1558

OY 263 KNASTPGLPGPKNISPPROVSVQKRSPPPKNVSPPP 299

DB 1559 qlvelpslppqpappqpqcpdpq--sspspsqlshlp 1594

RESULT 14  
 AAR63508

ID AAR63508 standard; Protein; 2842 AA.

AC AAR63508;

DT 23-MAY-1995 (first entry)

DE Adenomatous polyposis coli tumour repressor.

KM Adenomatous polyposis coli; tumour repressor; Gardner's syndrome;  
 KM familial adenomatous polyposis; cancer diagnosis and prognosis;  
 KM tumorigenesis suppression.

OS Homo sapiens.

PN US5352775-A.

PD 04-OCT-1994.

XX 08-AUG-1991; 91US-0741940.

PR 16-JAN-1991; 91GB-0000962.

PR 16-JAN-1991; 91GB-0000963.

PR 16-JAN-1991; 91GB-0000974.

PR 16-JAN-1991; 91GB-0000975.

PA (CANC-) CANCER INST.

PA (ICIL ) IMPERIAL CHEM IND.

PA (UYJO ) UNIV. JOHNS HOPKINS.

PA (UTAH ) UNIV UTAH.

PI Albersen H, Anand R, Carlson M, Groden J, Hedge PJ;

PI Joseph G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;

PI Vogelstein B, White RL.

DR MPI: 1994-316233/39.

DR N-PSDB; AAO72297.

XX New human adenomatous polyposis coli DNA encoding tumour

XX repressor - and derived primers and probes for diagnosis,

XX prognosis and treatment of cancer

XX Claim 3; Columns 71-84; 113pp; English.

XX AAO72297 is a cDNA isolated from the human adenomatous polyposis

XX coli (APC) gene, it encodes the tumour repressors described in AAR63507

XX CC and AAR63508. Determination of alterations in APC or its expression

XX CC products, can be used for the diagnosis and prognosis of cancer e.g.

XX CC colorectal, lung and breast tumours; and for determining predisposition

XX CC to certain cancers such as familial adenomatous polyposis (FAP) and

XX CC Gardner's syndrome. The wild type APC gene (or a part of it) can be

XX CC used therapeutically to restore gene function, while primers and probes

XX CC derived from the cDNA (AAO72333-400 and AAO72541-568) can be used to

XX CC detect mutations. Also APC proteins or analogues can be administered to

XX CC compensate for a defective gene, and epithelial cells, or transgenic

XX CC animals carrying a mutated APC allele are useful for detecting

XX CC therapeutic agents able to suppress tumorigenesis.

SO Sequence 2842 AA;

Query Match 5.9%; Score 103; DB 15; Length 2842;

Best Local Similarity 22.3%; Pred. No. 4.2;

Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

OY 79 PFLERQPKVYHRSQITLPRFSKHGVHRELSPTGSOQSRKRVVDYHSYVMOLESIRNE 138

DB 2151 pdgeekpfmskngpilkpkgekslekkiesekyikgkv--ysilttyvrs----- 2204

OY 139 GPPHLAGDRPSKHILKLVFTRHCIRALRLPRISIDLMSPLNT--SGEALSPYAKDE 196

DB 2205 --nseisgq-----mkxpiga-nmpsistsgrtmhlpgyvrnsstspvskgpp 2251

OY 197 ITOMILKSAARSELGMVSKNOEFYLRARRRRRRFAMKPVLOSISEMKVPMERHTPMAYR 256

```
Db 2252 ltkpaskspsegqtat-tspr-----gakpsvk--selspvar-qtsgq1g 2293
OY 257 DSGSPPKNA---STPSLPGPKNISPPOQVSPQSSPPPKNVSPPOPAFAVARTASKYSA 313
Db 2294 sskapsrsgrdstcprspaqqlsrplqspgrnslspgrnlsppnkslqlprtspscta 2353
OY 314 ASQOYORNRGNNAKSLY 329
Db 2354 st---ksqsgskmsy 2365
```

## RESULT 15

```
AA026052
ID AAR26052 standard; Protein: 2843 AA.
```

```
XX AC AAR26052;
```

```
XX DT 28-JAN-1993 (first entry)
```

```
XX DE APC gene product in familial adenomatous polyposis.
```

```
XX KW neoplasia; cancer; oncogene; tumour; growth; detection; diagnosis;
prognosis; treatment; sporadic colorectal carcinomas; ss.
```

```
XX OS Homo sapiens.
```

```
XX PN MO9213103-A.
```

```
XX PD 06-AUG-1992.
```

```
XX PE 16-JAN-1992: 92MO-US00376.
```

```
XX PR 16-JAN-1991: 91GB-0000963.
08-AUG-1991: 91US-0741940.
```

```
XX XX
```

```
PA (CANC-) CANCER INST.
```

```
PA (ICIL ) IMPERIAL CHEM IND PLC.
```

```
PA (UYJO ) UNIV JOHNS HOPKINS.
```

```
PA (UTAH ) UNIV UTAH.
```

```
XX XX
```

```
PI Albertsen H, Anand R, Carlson ML, Groden JL, Hedge PJ,
```

```
PI Joslyn J, Kinzler KW, Markham A, Nakamura Y, Thliveris A,
```

```
PI Vogelstein B, White RL, Markham AF;
```

```
XX DR WPI: 1992-284685/34.
```

```
XX DR N-PSDB: AAQ27234.
```

```
XX PT Detection of somatic and germ-line alterations of human APC gene
```

```
XX PT - used to diagnose, treat and study familial adenomatous
```

```
XX PS polyposis and sporadic colorectal cancer
```

```
XX PS Disclosure: Page 47; 132pp; English.
```

```
XX CC This sequence is encoded by the APC (Adenomatous Polyposis Coll)
```

```
XX CC gene associated with tumorigenesis, found on chromosome 5q.
```

```
XX CC The sequence may be mutated by deletions, insertions, or
```

```
XX CC point mutations of the gene. The APC gene is expressed in most normal
```

```
XX CC tissues as well suggesting that APC is a tumour suppressor.
```

```
XX SQ Sequence 2843 AA;
```

```
Db 2206 --nseisgq-----mkqplqa-nmpsisrgtmihpgvrnssstspvskkpp 2252
OY 197 ITOMILKSAARSELGMVYSKROEFYLRRARRRRKFAWKPVLOSISEMKPVMFHTPMAVR 256
Db 2253 ltkpaskspsegqtat-tspr-----gakpsvk--selspvar-qtsgq1g 2294
OY 257 DSGSPPKNA---STPSLPGPKNISPPOQVSPQSSPPPKNVSPPOPAFAVARTASKYSA 313
Db 2295 sskapsrsgrdstcprspaqqlsrplqspgrnslspgrnlsppnkslqlprtspscta 2354
OY 314 ASQOYORNRGNNAKSLY 329
Db 2355 st---ksqsgskmsy 2366
```

Search completed: September 16, 2002, 00:42:28  
Job time: 3377 sec



QY 233 MKPVLOSISEMKPVMEFHTPMAYRDSGSPKNA-----STPSLPGPKNISPPROVSVQ 286  
DB 482 MWPELSLSLDPPEARLPTDEVPESISLTQASPPVRAALDPOGSPPPDEPPAPRPVRLGPPPT 541  
QY 287 RSPSPKNA---VSPPOPAFVARTASKYSNAASQOVORANGNAK 326  
DB 542 KTLPTREGNLASPPPTLVGARTEETGTGPELSPAGPGESE 584

## RESULT 2

US-08-317-310A-64  
; Sequence 64, Application US/08317310A  
; Patent No. 5858701  
; GENERAL INFORMATION:  
; APPLICANT: WHITE, Morris F.  
; APPLICANT: SUN, Xiao Jian  
; APPLICANT: PIERCE, Jacalyn H.  
; TITLE OF INVENTION: THE IRS FAMILY OF GENES  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,310A  
; FILING DATE: 03-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Louis Myers  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: JDP-022  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1321 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-317-310A-64

Query Match 6.0%; Score 103.5; DB 2; Length 1321;  
Best Local Similarity 22.2%; Pred. No. 0.14;  
Matches 73; Conservative 30; Mismatches 93; Indels 133; Gaps 16;

QY 30 SCASVHDLGNDYVCGSTPYTTINMAAEVLEGVYFVCATPNAGPLEKQPKVYH 89  
DB 806 SCS-----GNDQYVLMSSP---VGRILEERLE-----PQATPGAGTP----- 841  
QY 90 RGSKILPRFSKHGVHRELRSPTHGSOQSRKVFYDHSVTMOQLISIRNEGPEPHLAGDRP 149  
DB 842 -----GAAGSGHTQ-----PHHSAVPS-----SMR-----PSAIGRP 869  
QY 150 SKHLKLVFTRHCLRALRLPRISID-----LMESPPLMLSGEALSPATAKDELITGMIL 202  
DB 870 EG-----FLGQCRKAVRPRLSLLEGLOTLPSSQDYPPLP---EPKSP----- 908  
QY 203 KSAARSELGMYVSKROEFLRRARRRRKFAKPKVLOSISEMKPVMEFHTPMAYRDSGSP 262  
DB 909 -----GEYIN---IDFGAGTRLSKPPAPPLASASSSSSLTASASPASSISGSGTPG 956  
QY 263 KNA-----STPSLPGPKN-----ISPPROVSVQORSPSPKNA 294

DB 957 TSSDSRORSPLSDYMNIDFSSPKSPKPTRSGDTVGSMDGLSP---EASSPPYPLPPRPS 1014  
QY 295 VSP-----PPOPAFVARTASKYSNAASQ 316  
DB 1015 TSPSSLOQPLPPAPGDLXRLPPASANTSQ 1043

## RESULT 3

US-07-741-940-7  
; Sequence 7, Application US/07741940  
; Patent No. 5352775  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARK  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner, Birch, McKie & Beckett  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/741,940  
; FILING DATE: 19920109  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107.035574  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2842 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: APC  
; US-07-741-940-7

Query Match 5.9%; Score 103; DB 1; Length 2842;  
Best Local Similarity 22.3%; Pred. No. 0.49;  
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PLEKOPRVHRSKILPRFSKHGVHRELRSPTHGSOQSRKVFYDHSVTMOQLSIRNE 138  
DB 2151 PDEKPKPTSNKGRILKPKCKSTLTETKIESESKIGKGRKV---YKSLTGVRS---- 2204  
QY 139 GPEHLADDRPSKHLKLVFTRHCLRALRLPRISIDLMESPPLNL---SGEALSPATAKOE 196

Db 2205 --NSEISGQ-----MKOPLQA-NMPSISRGRTMIHIGVNRSSSTSPVSKGPP 2251  
Oy 197 ITOMILKSAARSELGMVYKROEFLRARRRRKFMKPVLOISSEKMPVMEHPTPMAYR 256  
Db 2252 LKTPASKSPSEGGTAT-TSPR-----GAKPSVK--SELSPVAR-QTSQIGG 2293  
Oy 257 DSGSPPKNA---STPSLPGPKNISPPROVSPQRSSPPPKNVSPPOPAFVARTASKYSA 313  
Db 2294 SSKAPSRSGSRDSTSPRPAQOPLSRPIQSPGRNSISPGRNGLSPPNKLSQLPRTSSPSTA 2353  
Oy 314 ASOOVORNRGNKASLY 329  
Db 2354 ST---KSSGSGKMSY 2365

RESULT 4  
US-08-289-548A-7  
Sequence 7, Application US/08289548A  
Patent No. 5648212  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THILVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/289,548A  
FILING DATE: 12-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,46943  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2842 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-08-289-548A-7

Query Match 5.9%; Score 103; DB 1; Length 2842;  
Best Local Similarity 22.3%; Pred. No. 0.49;  
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

Oy 79 PFLEROPKYVHRGSKILFRFSKRGVHRELSPRTGSOQSRKVEDYHVTMOQLESINE 138  
Db 2151 PDEERKFTSNKNGPRILKPKCEKSTLETKKIESESGKIGGKKV--YKSLITGKVNS--- 2204  
Oy 139 GPEPHLAGDRPSKHLKLVFIHICLRALRLPRISIDLMESPLPNC--SGEALSPATYAKDE 196  
Db 2205 --NSEISGQ-----MKOPLQA-NMPSISRGRTMIHIGVNRSSSTSPVSKGPP 2251  
Oy 197 ITOMILKSAARSELGMVYKROEFLRARRRRKFMKPVLOISSEKMPVMEHPTPMAYR 256  
Db 2252 LKTPASKSPSEGGTAT-TSPR-----GAKPSVK--SELSPVAR-QTSQIGG 2293  
Oy 257 DSGSPPKNA---STPSLPGPKNISPPROVSPQRSSPPPKNVSPPOPAFVARTASKYSA 313  
Db 2294 SSKAPSRSGSRDSTSPRPAQOPLSRPIQSPGRNSISPGRNGLSPPNKLSQLPRTSSPSTA 2353  
Oy 314 ASOOVORNRGNKASLY 329  
Db 2354 ST---KSSGSGKMSY 2365

RESULT 5  
US-08-452-654-7  
Sequence 7, Application US/08452654  
Patent No. 5691454  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THILVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,654  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2842 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-08-452-654-7

Query Match 5.9%; Score 103; DB 1; Length 2842;  
Best Local Similarity 22.3%; Pred. No. 0.49;  
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PLEBPVKVHGSKILPREFSKGVHRELSPTHSGQSRKVDYHSVTMOQLSEIRNE 138  
DB 2151 PDEEKPFTSNKGRILKPEKSTLETKIESKIGKGV--YKSLITGKVR--- 2204  
QY 139 GPEHLAGDRPSKHUKLVETIRCLRALRPRISIDLESPLNL--SGEALSPYATKDE 196  
DB 2205 --NSEISGQ-----MKOPLQA-NMPSISRGRTMIHIPGVNRSSTSPVSKGPP 2251  
QY 197 ITOMILKSAARSELCMYVSKROEFLRARRRRKFAWKPVLOSISEMKPVMEFTPMAYR 256  
DB 2252 LKTPAKSPSEGOAT-TSPR-----GAKPSVK--SELSPVAR-QTSQIGG 2293  
QY 257 DSGSPKNA---STPSLPGKNTISPPROYVPORSSPPKNSVPPQAFVARTASXYSA 313  
DB 2294 SSKAPRSRSGSDSTSPRAQOPLSRPQSPGRNSISPGRNGISPPKLSQLPRTSSPSTA 2353  
QY 314 ASQOVORNRGNNAKSLY 329  
DB 2354 ST---KSSGSGKMSY 2365

RESULT 6  
US-07-741-940-2  
Sequence 2, Application US/07741940  
Patent No. 5352775  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07741,940  
FILING DATE: 19920109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9289

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-741-940-2

Query Match 5.9%; Score 103; DB 1; Length 2843;  
Best Local Similarity 22.3%; Pred. No. 0.49;  
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PLEBPVKVHGSKILPREFSKGVHRELSPTHSGQSRKVDYHSVTMOQLSEIRNE 138  
DB 2152 PDEEKPFTSNKGRILKPEKSTLETKIESKIGKGV--YKSLITGKVR--- 2205  
QY 139 GPEHLAGDRPSKHUKLVETIRCLRALRPRISIDLESPLNL--SGEALSPYATKDE 196  
DB 2206 --NSEISGQ-----MKOPLQA-NMPSISRGRTMIHIPGVNRSSTSPVSKGPP 2252  
QY 197 ITOMILKSAARSELCMYVSKROEFLRARRRRKFAWKPVLOSISEMKPVMEFTPMAYR 256  
DB 2253 LKTPAKSPSEGOAT-TSPR-----GAKPSVK--SELSPVAR-QTSQIGG 2294  
QY 257 DSGSPKNA---STPSLPGKNTISPPROYVPORSSPPKNSVPPQAFVARTASXYSA 313  
DB 2295 SSKAPRSRSGSDSTSPRAQOPLSRPQSPGRNSISPGRNGISPPKLSQLPRTSSPSTA 2354  
QY 314 ASQOVORNRGNNAKSLY 329  
DB 2355 ST---KSSGSGKMSY 2366

RESULT 7  
US-08-289-548A-2  
Sequence 2, Application US/08289548A  
Patent No. 5648212  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/289,548A  
FILING DATE: 12-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.46943  
TELECOMMUNICATION INFORMATION:







1 ADDRESS: BANNER & MITCOFF, LTD.  
2 STREET: 1001 G STREET, N.W.  
3 CITY: WASHINGTON  
4 STATE: DC  
5 COUNTRY: US  
6 ZIP: 20001  
7  
8 COMPUTER READABLE FORM:  
9 MEDIUM TYPE: Floppy disk  
10 COMPUTER: IBM PC compatible  
11 OPERATING SYSTEM: PC-DOS/MS-DOS  
12 SOFTWARE: Patent In Release #1.0, Version #1.25  
13 CURRENT APPLICATION DATA:  
14 APPLICATION NUMBER: US/08/370,235A  
15 FILING DATE: 01-JAN-1995  
16 CLASSIFICATION: 435  
17  
18 ATTORNEY/AGENT INFORMATION:  
19 NAME: KAGAN, SARAH A.  
20 REGISTRATION NUMBER: 32,141  
21 REFERENCE/DOCKET NUMBER: 01107.48688  
22 TELECOMMUNICATION INFORMATION:  
23 TELEPHONE: 202 508 9100  
24 TELEFAX: 202 508 9299  
25 INFORMATION FOR SEQ ID NO: 2:  
26 SEQUENCE CHARACTERISTICS:  
27 LENGTH: 2843 amino acids  
28 TYPE: amino acid  
29 TOPOLOGY: linear  
30 MOLECULE TYPE: protein  
31 US-08-370-235A-2

```

Query Match Similarity      5.9%: Score 103; DB 2; Length 2843:
Best local similarity      22.3%: Pred. No. 0.49;
Matches      57; Conservative      46; Mismatches      107; Indels      46; Gaps      11.

QY      79  PLETERPKVYHRSKILLPFSKHGVHRELSPHSGOOSRKVFDYHVTYMOLESIRNE 138
           + : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2152 PDGEKPEFPMNSGRILKPGKSTLETKEIESGKIGKGV--YKSLITGVKS---- 2205

QY      139 GPEPHIADGRSKHLKLVFTRHCLRALRLPRISIDLMESLPLM--SGALSPLATAKDE 196
           + : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2206 ---NSEISG-----WKQPLQAHNPSTISGRRTMHPGVNRNSSSTSPVSKGPP 2252

QY      197 ITQMILKSANSELGMVYSKROEFLYLRARRRRKFAWKPYLOSISBMKVMFEHTPMAYR 256
           + : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2253 LKTPASKSPSGQFAT--TSPR-----GAKPSVK--SELSPVAN-QTSGIQG 2294

QY      257 DSGSPPKNA-----STPSLPGPKNISPPROYVYPQSSSPPPKNVAPPQAPFAKTAASKYA 313
           + : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2295 SSAPSRGSDSDSPSPKPAQOPLSPRIQSPGRNISPGRNGISPPNKLSQLPSTSPSTA 2354

QY      314 ASQOQVORNGNAKSLY 329
           + : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2355 ST---KSSGSGKMSY 2366

RESULT 12
US-08-450-582-2
: Sequence 2, Application US/08450582
: Patent No. 6114124
: GENERAL INFORMATION:
: APPLICANT: ALBERTSEN, HANS
: APPLICANT: ANAND, RAKESH
: APPLICANT: CARLSON, MARY
: APPLICANT: GRODEN, JOANNA
: APPLICANT: HEDGE, PHILIP J.
: APPLICANT: JOSLYN, GEOFF
: APPLICANT: KINZLER, KENNETH
: APPLICANT: MARKHAM, ALEXANDER F.
: APPLICANT: NAKAMURA, YUSUKE
: APPLICANT: THLIVERIS, ANDREW
: TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
: TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

```

```

1      NUMBER OF SEQUENCES: 102
2      CORRESPONDENCE ADDRESS:
3      ADDRESSEE: Banner & Witcoff, Ltd.
4      STREET: 1001 G Street, NW
5      CITY: Washington
6      STATE: D.C.
7      COUNTRY: USA
8      ZIP: 20001-4598
9      COMPUTER READABLE FORM:
10     MEDIUM TYPE: Floppy disk
11     COMPUTER: IBM PC compatible
12     OPERATING SYSTEM: PC-DOS/MS-DOS
13     SOFTWARE: PatentIn Release #1.0, Version #1.25
14     CURRENT APPLICATION DATA:
15     APPLICATION NUMBER: US/08/450,582
16     FILING DATE:
17     CLASSIFICATION: 435
18     PRIOR APPLICATION DATA:
19     APPLICATION NUMBER: US 08/452,655
20     FILING DATE: 25-MAY-1995
21     APPLICATION NUMBER: US 08/289,548
22     FILING DATE: 12-AUG-1994
23     PRIOR APPLICATION DATA:
24     APPLICATION NUMBER: US 07/741,940
25     FILING DATE: 08-AUG-1991
26     ATTORNEY/AGENT INFORMATION:
27     NAME: Kagan, Sarah A.
28     REGISTRATION NUMBER: 32,141
29     REFERENCE/DOCKET NUMBER: 1107.49964
30     TELECOMMUNICATION INFORMATION:
31     TELEPHONE: 202-508-9100
32     TELEFAX: 202-508-9299
33     INFORMATION FOR SEQ ID NO: 2:
34     SEQUENCE CHARACTERISTICS:
35     LENGTH: 2843 amino acids
36     TYPE: amino acid
37     TOPOLOGY: linear
38     MOLECULE TYPE: protein
39     US-08-450-582-2

```

```

Query Match Similarity      5.9%: Score 103; DB 3; Length 2843:
Best Local Similarity      22.3%; Pred No. 0.49;
Matches 57; Conservative   46; Mismatches 107; Indels 46; Gaps 11.

OY      79 PLEERPKVHVGSKILPRFSKHGVNRELSPTHSGSOASKVEDHYSVTMOOLESTINE 138
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db       2152 PDDEEFPFNSNGKPIIKPGKSTLLETKKIESESKIGGCKV--YKSLLTGKVAS---- 2205

OY      139 GPBPHLAGRPSKHLKLVETRHCTALRLPRISIDLMESPJPN--SGEALSPTATKDE 156
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db       2206 --NSELSIQ-----MKQPLQA-NMPSISKRRTMIHPGVRRSSSSTIPYSKKGP 2252

OY      197 ITOMILKSAARSELGMVYVKROEFYLRAARRRRRKFMKPVLQGISIMKPEMEHTPMAYR 256
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db       2253 LKRPASKSPSEGOAT-TSPR-----GAKPSYK--SELISVAR-QTSOTIG 2294

OY      257 DSGSPKNK---SPFLSGPKNISPPROYVYPQSRSSPKNKVPPOPAFYVTRATSKYA 313
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db       2295 SSNAFRSSRSRSTSPRAPAOPLSRFIQSPGNISIFORNGISPPKLSQLPRTISSPSTA 2354

OY      314 ASQQOVORNRGNKASLY 329
          | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db       2355 ST---KSSGSGCKMSY 2366

RESULT 13
US-08-582-7
; Sequence 7, Application US/08450562
; Patent No. 6114124
; GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
;
APPLICANT: ANAND, RAKESH
;
```





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2002, 00:34:21 ; Search time 42.86 Seconds  
(without alignments)  
742.080 Million cell updates/sec

Title: US-09-828-313-35

Perfect score: 1734  
Sequence: 1 MGITPSCVTVQGVYRVYRP.....SAASQOVRNKNKSLYMA 331

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089314 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131	7.6	413	2	112533
2	124	7.2	796	2	E96654
3	118.5	6.8	1102	2	T28666
4	117	6.7	891	2	T40503
5	117	6.7	891	2	A38903
6	116	6.7	912	2	A54423
7	114	6.6	742	2	P84643
8	112.5	6.5	576	2	T38293
9	111.5	6.4	539	2	T33658
10	111	6.4	651	2	T01510
11	109	6.3	651	2	T16754
12	108	6.2	1356	2	T05441
13	107	6.2	379	2	D85257
14	107	6.2	603	2	P66797
15	107	6.2	603	2	T41643
16	107	6.2	1461	2	T18860
17	105.5	6.1	830	2	T18860
18	105	6.1	545	2	T33659
19	105	6.1	568	2	T33660
20	105	6.1	2845	2	I49505
21	104.5	6.0	377	2	A48018
22	104	6.0	597	2	AH2351
23	104	6.0	1151	2	T18535
24	104	6.0	2254	2	D86215
25	103.5	6.0	2022	2	T48818
26	103	5.9	2783	1	A41948
27	103	5.9	2843	1	R8H0AP
28	102.5	5.9	394	2	C84905
29	102	5.9	620	2	S06733

30	102	5.9	1217	2	T39427	probable myosin I
31	101.5	5.9	653	2	E84662	hypothetical prote
32	101.5	5.9	1201	2	G86441	unknown protein (l
33	100.5	5.8	530	2	G70904	hypothetical prote
34	100	5.8	561	2	T16148	hypothetical prote
35	100	5.8	645	2	A71416	hypothetical prote
36	100	5.8	678	2	JC4245	transcription fact
37	100	5.8	734	2	T04876	hypothetical prote
38	99.5	5.7	479	2	D71809	hypothetical prote
39	99.5	5.7	1183	2	T39223	probable inositol
40	99	5.7	464	2	S22697	extensin - Volvox
41	99	5.7	713	2	B32575	C-sk1 protein PB28
42	99	5.7	847	2	F96531	hypothetical prote
43	99	5.7	990	2	T12678	hypothetical prote
44	99	5.7	1021	2	T15765	hypothetical prote
45	99	5.7	3530	2	A59266	unconventional myo

ALIGNMENTS

RESULT 1  
T12533  
hypothetical protein DKFP434B034.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 23-Jul-1999  
R:Wamult, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
A:Accession: T12533  
A:Reference number: Z17524  
A:Submitted to the Protein Sequence Database, June 1999  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residue type: 1-413 <MANE>  
A:Cross-references: EMBL:AL096728  
A:Experimental source: adult testis; clone DKFP434B034  
C:Genetics:  
A>Note: DKFP434B034.1

Query Match 7.6%; Score 131; DB 2; Length 413;

Best local similarity 20.6%; Pred. No. 0.017; Matches 78; Conservative 57; Mismatches 142; Indels 102; Gaps 15;

QY	3	LPFPSCVTVQGVYRVYRP--DGHENLSKSGSVH-----DLIDNPYYVCGSTPTITN 55	
DB	42	MPSPNMAIVLG-PNLLMFQADGNTTMTTVSLQIVGIIPEPIQHADMFFRGELERNING 100	
QY	56	RNAAEVLEYGTVFVCATPNAOFLEORPKVYHR-----GSKILPFSKHGVHYRELR 109	
DB	101	NYGSPVHVNNHNNANYSMPSPMDPADRPEQARRPLSVATDNMKLEFYKDKDMGVYMD 160	
QY	110	S---PTHGQSQRKVFEDVHVTMOQLESIRNEGPEPHLAGDRPSKHILKIVFIRHCLRALR 166	
DB	161	TNNVARRGSSAKRY-----SCAPPSMQPRAPRAELDALPLSP----- 198	
QY	167	LPRIIDLESPLPULSGEALSP---TATAKDEITQMLTKAAR----- 207	
DB	199	LPQPLDSPAALSPSGIGLQPGPERTSTTKSK--ELSPGSAQSGSPSSQGTACAGTQ 256	
QY	208	-----SEIGMYVSKROEFYLLRRARRRRKFAWPKVLOSISEKKPYMEFHP--MAYRDSG 259	
DB	257	PGAOPGAQAGASPSQPPRADOHPHLTK-----VSKKLAPRPVPRPGQGANADQASG 311	
QY	260	SPPK---NASTPSLGPKNISPPQVSP-----QSSPPPK- 293	
DB	312	QPSVSLSPPTSPPTSPYGLSTYPOGYSLASGOLSPAADPLASPSVFTTSLSKRPTPKR 371	
QY	294	---NVSPPOPAFVARTAS 309	
DB	372	KQRPILPPQPPPTVNLAS 390	

## RESULT 2

hypothetical protein F16P17.12 [imported] - Arabidopsis thaliana  
E96654  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E96654  
R:Theologian, A.; Ecker, J.R.; Palm, C.T.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 815-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Martzball,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: E96654  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1796 <STO>  
A:Cross-references: GB:AE005173; NID:g8493585; PIDN:AAE75808.1; GSPDB:GN00141  
A:Gene: F16P17.12  
A:GeneID:5  
A:Map position: 1

Query Match 7.2%; Score 124; DB 2; Length 796;  
Best Local Similarity 23.6%; Pred. No. 0.14;  
Matches 68; Conservative 36; Mismatches 132; Indels 52; Gaps 11;

74 TPNAG-----PLERPPKVVHGSKILPRSKHGIVHRELRSPTGSOQSRK 120  
164 TPOSOSQPHHQOOSQSPQHPNQNQVSPQNOQLF-FQLQHPFAQAVOTQSOQSOQRS 222  
121 VFVHNVHTMOQLESIRNGRPHPLAGD-----RPSKHLKVFIRHCLRALRPRISIDM 175  
223 QFDOQLLEQOQASRQOVQEHSSGSSGDMKRPVEVELINLNAAERQVSSRRAREPE 282  
176 ESPPLNLSGELSPATAKDEIT-----QKILKSARSEL-CMVYSROEFLYLRARRR 228  
283 RPPRSRREG---SHRTPSTELTMAKRPVSPVSHSLVWQVSEPAQOYL--SSRS 337  
229 RKRMRKPVQSISEKKPVMEHTPMAYRDSG---SPKNASTSTELGPKNTSPRPQV 284  
338 SEAAQLSLLEPSVSDSS-----HASQPTRSNQSHAVSKPQVPVSKPMPMSQPPP--TSN 390  
285 PORSPPPKVNSP-----PPQPAFYARTASKYSAASQOVQRNR 322  
391 PFLSQPPSKSKPFPMSQSONSKRPFVSSQSKRPLLVSSQSQRSK 438  
DB

## RESULT 3

728666  
protein kinase C-related kinase PRKSD - Suberites domuncula  
C:Species: Suberites domuncula  
C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 15-Sep-2000  
C:Accession: T28666  
R:Mueller, W.E.G.  
submitted to the EMBL Data Library, April 1997  
A:Reference number: 220501  
A:Accession: T28666  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 11102 <MD>  
A:Cross-references: EMBL:Y13104; NID:e1289674; PID:e316788; PIDN:CAA73558.1  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 6.8%; Score 118.5; DB 2; Length 1102;  
Best Local Similarity 25.8%; Pred. No. 0.54;  
Matches 85; Conservative 36; Mismatches 149; Indels 59; Gaps 18;

OY 36 LLGPNPOYVCGSTPYTTTNRMAAEVLEYGVTFYCATPMAOPFLEROPKVVHGSKTL 95

DB 451 LVRLRDLYPDGTGATYCLP--LEPQGLILEVYTFEPKTERKOPKRLRGRIIRRG-KVL 507

OY 96 PRESKHGIVHRELRSPTGSOQSRKVEDYNS-----VTMOQLESIRNGP--EHP 143

DB 508 ---RH---TELNTDVTMARLLNRSAPSTHTDPHSSPWNVISVQKSSKQSEPPSPH 559

OY 144 LAG---DBPSKHLKVFIRHCLRALRPRISIDLMESPLNLSGELSPATAKDEITQ 199

DB 560 TSGSTPKSPKIPPTSPYFPHDSGRKSLDENGDL-PPPIP-KSEEDL--IVAPHSATE 615

OY 200 MIKSAARSELGMVYSKRQEFYLRARRRRKFAKRPVLOSISEKKPVMEHTPMAY--RD 257

DB 616 PII-PPPRFLGVSSTDE---HRVDNIGPPPKFIMESWAVSPTHRTTSPOHQKPRD 671

OY 258 SG-----SPPKNASTRPS-----LPGPKNTSPRPQVSPORSPPKSPPPQ 300

DB 672 TGREYMTVISPPTSPVPAPTSPANQVPPRPNSPPTPLS-PSRVADLPDISQPPPA 730

OY 301 PAFVARTASKYSAASQOVQRNRGNASKLY 329

DB 731 HTRVNRRSIEGS---QMPQVYGNVNF 755

## RESULT 4

740503  
protein kinase kin1 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000

C:Accession: T40503  
R:Gwilliam, R.; Rajadream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.  
submitted to the EMBL Data Library, September 1998

A:Reference number: 221933  
A:Accession: T40503  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-891 <GMT>

A:Cross-references: EMBL:AL031534; PIDN:CAA20726.1; GSPDB:GN00067; SPDB:SPBC4F6.06

A:Experimental source: strain 972h-; cosmid c4f6

A:Gene: SPDB:SPBC4F6.06

A:Map position: 2

A:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

Query Match 6.7%; Score 117; DB 2; Length 891;  
Best Local Similarity 21.8%; Pred. No. 0.53;

Matches 78; Conservative 50; Mismatches 137; Indels 92; Gaps 17;

OY 17 VVYPDGHVEN-LSKSCVNDLLGNPDY-----VCGSTPYTTTNRMAE----- 60

DB 262 VVHRDLKINILISKTDGIKIIDGLSNLYRQSRILRTFCGLYFAPELLNNQPIYIGE 321

OY 61 -EVLEYGVTFYCATPMAOPFLERQPKVH---RQSKILPRESKHGIVHRELRSPTG 115

DB 322 VDWMSGIYLVYVGVKV-PPDDNMCALAKIKGTVEYPSY-----LSSQCKL 371

OY 116 QOSKRVFD-YHSTVMOQLES---IRN-EGPEPHLAGDRPSKHLKVFIRHCLRALR 169

DB 372 LSRMLVTDPLKNTLLEVLNHPMIMRYEGPPASFAPEKSP----- 412

OY 170 ISIDLMESPLNLSGELSPATAKDEITOMILKSARS--ELGW-----VSKROEFL 222

DB 413 ITLPLDPEITREKNGFDGPEKIVRELTKVISSEHAYOSLAKTGFYSGPMSADKASFPE 472

OY 223 RRARRRRKFAKRPVLOSISEKKPVMEHTPM-----AYRDSGSPPKNASTRP--S 269

DB 473 FRIRHAHNDIENPILPISLNTDIDYDAFHLISYIYLVSEERYVEKGMNRRITAKTPVSS 532

OY 270 LPCKNTSPRPQVSPORSPPKVPPOPAFVARTASKYSAASQOVQRNRGNK 326

DB 533 VP-----SSPVO-----PTSYNRTLPMPPEV---AYKQDESPRVSRVNRNLSLAR 573



RESULT 5  
A: Molecule type: mRNA  
A: Residues: 1-912 <YAM>  
A: Cross-references: GB:M64999; NID:9173409; PIDN:AAA63577.1; PID:9173410  
C: Superfamily: Aggreacan; C-type lectin homology; complement factor H repeat homology,  
F:50-139/Domain: immunoglobulin homology <IMM>  
F:174-251/Domain: link protein repeat homology <LNK1>  
F:272-353/Domain: link protein repeat homology <LNK2>  
F:651-682/Domain: EGF homology <EGF>  
F:689-809/Domain: C-type lectin homology <LCH>  
F:816-872/Domain: complement factor H repeat homology <FHD>

Query Match 6.7%; Score 117; DB 2; Length 891;  
Best Local Similarity 21.8%; Pred. No. 0.53;  
Matches 78; Conservative 50; Mismatches 137; Indels 92; Gaps 17;

QY 17 VYRPGHVEN-LKSCSVDDLGNPDY-----VCGSTYTTTNMAAE----- 60  
DB 262 VVHRDLKLEILISKTGDIKIDFGSLNLYRQSLRTFCCLFPAPELLNAPYIGPE 321  
QY 61 -EVLEYGVTVFCATPNAOPFLERQPKVVH-----RGSKILPRFSKHGVHRELSPTGHS 115  
DB 322 VVWVSGFGLVVLVCGKV-PFDQNNKSAHAKIKKGYEYSY-----LSSCKKL 371  
QY 116 QOSKRVFP-YHSVTMOQLS-----INN-EGPEPHLAGRPSKHLKLVTRHCLRALR 169  
DB 372 LSRMLVTPLKRALTELEVNLNHPMMIRNYEGPPASFPAPERSP----- 412  
QY 170 ISIDMESPLNLSGEALSPATAKDEITOMILKSAARS--ELGMY-----VSKROEFL 222  
DB 413 IITLPDPELTENMGDFGPPREKTIVRELTKVLSSEAYQSLAKTGFYSGPNSADKKSFTE 472  
QY 223 RARRRRRFAMKPVLOSISEMKFVMEFHTPM-----AYRDSGSPKNASTP--S 269  
DB 473 FIRIRAHADIENPILPSLISMTDIDAFHPLISYLVSESRVYKEKGNMNRIRAKTPVSS 532  
QY 270 LKGPKNISPPROVSVPPQSSPPKNAVSPPOPAFAVARTASKYSAASQOVORNGNAK 326  
DB 533 VP-----SSPQV-----PTSYNRTLPMPBYV---AYGDESESRVSKNTSLAK 573

RESULT 6  
A: Molecule type: mRNA  
A: Residues: 1-912 <YAM>  
A: Cross-references: GB:M64999; NID:9173409; PIDN:AAA63577.1; PID:9173410  
C: Superfamily: Aggreacan; C-type lectin homology; complement factor H repeat homology,  
F:50-139/Domain: immunoglobulin homology <IMM>  
F:174-251/Domain: link protein repeat homology <LNK1>  
F:272-353/Domain: link protein repeat homology <LNK2>  
F:651-682/Domain: EGF homology <EGF>  
F:689-809/Domain: C-type lectin homology <LCH>  
F:816-872/Domain: complement factor H repeat homology <FHD>

Query Match 6.7%; Score 117; DB 2; Length 891;  
Best Local Similarity 21.8%; Pred. No. 0.53;  
Matches 78; Conservative 50; Mismatches 137; Indels 92; Gaps 17;

QY 17 VYRPGHVEN-LKSCSVDDLGNPDY-----VCGSTYTTTNMAAE----- 60  
DB 262 VVHRDLKLEILISKTGDIKIDFGSLNLYRQSLRTFCCLFPAPELLNAPYIGPE 321  
QY 61 -EVLEYGVTVFCATPNAOPFLERQPKVVH-----RGSKILPRFSKHGVHRELSPTGHS 115  
DB 322 VVWVSGFGLVVLVCGKV-PFDQNNKSAHAKIKKGYEYSY-----LSSCKKL 371  
QY 116 QOSKRVFP-YHSVTMOQLS-----INN-EGPEPHLAGRPSKHLKLVTRHCLRALR 169  
DB 372 LSRMLVTPLKRALTELEVNLNHPMMIRNYEGPPASFPAPERSP----- 412  
QY 170 ISIDMESPLNLSGEALSPATAKDEITOMILKSAARS--ELGMY-----VSKROEFL 222  
DB 413 IITLPDPELTENMGDFGPPREKTIVRELTKVLSSEAYQSLAKTGFYSGPNSADKKSFTE 472  
QY 223 RARRRRRFAMKPVLOSISEMKFVMEFHTPM-----AYRDSGSPKNASTP--S 269  
DB 473 FIRIRAHADIENPILPSLISMTDIDAFHPLISYLVSESRVYKEKGNMNRIRAKTPVSS 532  
QY 270 LKGPKNISPPROVSVPPQSSPPKNAVSPPOPAFAVARTASKYSAASQOVORNGNAK 326  
DB 533 VP-----SSPQV-----PTSYNRTLPMPBYV---AYGDESESRVSKNTSLAK 573

A: Molecule type: mRNA  
A: Residues: 1-912 <YAM>  
A: Cross-references: GB:M64999; NID:9452820; PIDN:CAA53481.1; PID:9452821  
C: Superfamily: Aggreacan; C-type lectin homology; complement factor H repeat homology,  
F:50-139/Domain: immunoglobulin homology <IMM>  
F:174-251/Domain: link protein repeat homology <LNK1>  
F:272-353/Domain: link protein repeat homology <LNK2>  
F:651-682/Domain: EGF homology <EGF>  
F:689-809/Domain: C-type lectin homology <LCH>  
F:816-872/Domain: complement factor H repeat homology <FHD>

Query Match 6.7%; Score 116; DB 2; Length 912;  
Best Local Similarity 24.2%; Pred. No. 0.66;  
Matches 54; Conservative 26; Mismatches 101; Indels 42; Gaps 8;

QY 140 PEPLIADRRSKILKLVFIRHCLRALRPR-----ISIDME-----GPLNLS 183  
DB 368 PASHIADALE--AIVVTELTLEELKLPQAVESRGAIVSPTIEDGGGGSSTPDP 424  
QY 184 GEALSPYATKDEITOMILKSAARSELGMVYKROEFLRARRRRK-----FA 232  
DB 425 AEA--PRTLEFE-TOSIVPPLIGSSEEGKVLQEEKYRGEKEEKEEVEDEALMA 481  
QY 233 WKPVLOSISEMKFVMEFHTPMAYRDSGSPKNA-----STPSLPGPKNISPPROVSV 266  
DB 482 WPSLSLDEADLPTEVPEVPSLTQASBPYRALOPGVSPPYDEPAEPKPVLCPT 541  
QY 287 RSSPPKN--VSPPOPAFAVARTASKYSAASQOVORNGNAK 326  
DB 542 KTLPTREGMLASPPSPSTLVGAREIEETGPELSCAPRGESE 584

RESULT 7  
A: Molecule type: mRNA  
A: Residues: 1-912 <YAM>  
A: Cross-references: GB:M64999; NID:9452820; PIDN:CAA53481.1; PID:9452821  
C: Superfamily: Aggreacan; C-type lectin homology; complement factor H repeat homology,  
F:50-139/Domain: immunoglobulin homology <IMM>  
F:174-251/Domain: link protein repeat homology <LNK1>  
F:272-353/Domain: link protein repeat homology <LNK2>  
F:651-682/Domain: EGF homology <EGF>  
F:689-809/Domain: C-type lectin homology <LCH>  
F:816-872/Domain: complement factor H repeat homology <FHD>

Query Match 6.6%; Score 114; DB 2; Length 742;  
Best Local Similarity 21.7%; Pred. No. 0.72;  
Matches 60; Conservative 48; Mismatches 127; Indels 42; Gaps 9;

QY 52 TITRMAEEVLEYGVTVFCATPNAOPFLERQPKVVHSGSKILPRFSKHGVHRELSR 111  
DB 352 TVFNQITANMIIDESLDSGSPRSPDSRLSLEALEKYEKTKIM-----ISENIYSSP 404  
QY 112 THGSQOSKRVFEDVHVTMOQLSIRNEGPEPHLAGRPSKHLKLVFIRHCLRALRPRIS 171  
DB 405 DTSSPEKE-----DTMSHSYAD--PNSILKKVDESGRLVSVQRVHNSKIFSPRMV 456  
QY 172 IDLESPLPULSGEALSPYATKDEITOMILKSAARSELGM-----YVSKROEFLRRA 225  
DB 457 QSPVSTPLPNSRSPTOGSPASISRFH-----SSPSLGLITSILHDHOSCKDESTSSSS 508  
QY 226 RRRRRFAMKPVLOSISEMKFVMEF-----HTPMAYRDSGSPKNASTPSPKPNISPPRO 281

Db 509 PASRSISFLPTLHPLTSSOPKKASPOCPSTPVHNSGPPSAEAAVTSSLPDP-LKPLRI 567  
 QY 282 VSVPRQSSPPPKNVSPPOPAFAVARTASKYSAASQY 318  
 Db 568 LSRP-----PPPPISLSLRSPSPSTNSI 596

## RESULT 8

hypothetical serine-rich protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T38293  
 R:Skellton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 submitted to the EMBL Data Library, January 1996  
 A:Reference number: 221784  
 A:Accession: T38293  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-576 <SKE>  
 A:Cross-references: EMBL:268887; PIDN:CAA93115.2; GSPDB:GNO00066; SPDB:SPAC23E2.03c  
 A:Experimental source: strain 972h-; cosmid c23E2  
 C:Genetics:  
 A:Gene: SPDB:SPAC23E2.03c  
 A:Map position: 1  
 A:Introns: 553/2

Query Match 6.5%; Score 112.5; DB 2; Length 576;  
 Best Local Similarity 23.8%; Pred. No. 0.68;  
 Matches 75; Conservative 35; Mismatches 120; Indels 85; Gaps 14;

QY 67 VTYFVCATPAQPFLEQRQKVVHNSGSKTLRFKSHCV-HVRELASPTGSOQSKRVYD 125  
 Db 18 ITGKICETNA-----ATSPRSFOQLRIGYAVYHGHFLSPSHKHPPTTASASDSD 69  
 QY 126 SVTMOLESIRNGEPRLAGDRPSKHL-----KLVTIRHCLR 163  
 Db 70 EAT-----EIFTHSQPLPIFAGPSSKSHADIFKFKFSKASSLSPCKSDSNMVICYMLK 125  
 QY 164 ALRLPRISI-----DLMEPLPLNLSGFLSP-----ATRAK-----DETTQILK 203  
 Db 126 ATISKRAFLGSSQSKAKELINVP-PNLACKRPLSANSFGSSSKTFQLPREDSTV-----P 180  
 QY 204 SAARSELGMVSKROEFYLRARRRRKFAKPVLOSTISEMKPYMEHTPP-----AYR 256  
 Db 181 SASLYKOPSFNSNPARTTSSATHTSQF-----STSSSSSVNSVHTTPVWVPVYFOYN 233  
 QY 257 DSGSPPKNASTPFLGPKNISPPQVSVPOKSSPP--PKNVSPPOPAFAVARTASKYSA 314  
 Db 234 SSMVAPSSSSSSVAP-----FVPRKQFVSSASDPQPTPTISMSPPIPP-----TESQFSAP 284  
 QY 315 SQOYQNRKGNANSY 329  
 Db 285 MYOQOQSYLSPOSHY 299

## RESULT 9

hypothetical protein W07B3.2c - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T33658  
 R:Wohlmann, P.; Bauer, C.; Rohlfing, T.; Giliam, B.  
 submitted to the EMBL Data Library, October 1998  
 A:Description: The sequence of C. elegans cosmid W07B3.  
 A:Reference number: 221384  
 A:Accession: T33658  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-539 <MOH>  
 A:Cross-references: EMBL:AF100304; PIDN:AAC68911.1; GSPDB:GNO0021; CESP:W07B3.2c  
 A:Experimental source: strain Bristol N2; clone W07B3

C:Genetics:  
 A:Gene: CESP:W07B3.2c  
 A:Map position: 3  
 A:Introns: 61/3; 101/3; 124/3; 257/3; 291/2; 364/3; 495/3

Query Match 6.4%; Score 111.5; DB 2; Length 539;  
 Best Local Similarity 23.0%; Pred. No. 0.74;  
 Matches 43; Conservative 38; Mismatches 75; Indels 31; Gaps 6;

QY 154 KLVTIRHCLRRLPRISID-LMEPLPLNLSGFLSPPTATKDEITOMILKSAARSELGM 212  
 Db 300 KRYVNRORTEAFRRRRREEEMLAMPGRINGEAL-----DRAQIVVRNKRAREAR 352  
 QY 213 VYSKQOEYTLRARRRKFAKPVLOSISEMKPYMEHTPPMAVYDSCSPPKNSPTSLPG 272  
 Db 353 LRYQMTPDQRKSTNOKRYTPKRRNEMDMISAGO-----NAVGSSTTMTGTTPAKKD 408  
 QY 273 P-----KNISPPQVSVPOKSSPPPKNVSPPOPAFAVARTASKYSAASQ 317  
 Db 409 DMDALSLEPDRVVKRTOHAQOILLRQANQSSGTPRSVTPPPQAYIISNGOQH-----QQ 464  
 QY 318 YQNRGN 324  
 Db 465 LVGNQGN 471

## RESULT 10

T01510  
 hypothetical protein T10M13.10 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 04-Mar-2000  
 C:Accession: T01510  
 R:Johnson, A.F.; de la Bastide, M.; Lodhi, M.; Hoffman, J.; Hasegawa, A.; Gnoj, L.; G.  
 Martienssen, R.; McCombie, W.  
 submitted to the EMBL Data Library, May 1997  
 A:Description: The sequence of the Arabidopsis thaliana T10M13 BAC.  
 A:Reference number: 214346  
 A:Accession: T01510  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-202 <JOH>  
 A:Cross-references: EMBL:AF001308; NID:g2104523; PID:g2104533  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 45  
 A:Note: T10M13.10  
 C:Superfamily: Arabidopsis thaliana hypothetical protein T10M13.10

Query Match 6.4%; Score 111; DB 2; Length 202;  
 Best Local Similarity 21.9%; Pred. No. 0.23;  
 Matches 54; Conservative 24; Mismatches 89; Indels 80; Gaps 7;

QY 16 RVVYPDGHVENLSKSCSVHDLIGNPDYVCGSTP-----YTTNRMA---AEVL 63  
 Db 19 KYVLSDBGVQNLBEETTVAAIEMLENQHVVEFDPSISFNDAKTKYKRLAFLPADKTL 78  
 QY 64 EYGVYFVCATPAQPFLEQRQKVVHNSGSKTLRFKSHCV-HVRELASPTGSOQSKRVYD 123  
 Db 79 EFGKILYV-----LPA-KRSGGRAAASSAVLTJSEMKML- 113  
 QY 124 YHSVYMOLESIRNGEPRLAGDRPSKHLKLVTIRHCLRRLPRISIDLMESPLPLNS 183  
 Db 114 -----FSATLAVNRSFSYEGILPMT 135  
 QY 184 GEALSPYATADETOMILKSAARSELGMVYSKROEFLRARRRKFAKPVLOSISM 243  
 Db 136 TRSYNNNNPARD--TVVAATSVGRLEAEEMEEDREFLSRQSG--GWKPSLPDIRK 190  
 QY 244 KPMEFH 250  
 Db 191 KAKKKIH 197



Query Match	6.28;	Score 107;	DB 2;	Length 379;
Best Local Similarity	48.98;	Pred. No. 1;		
Matches 22; Conservative	6;	Mismatches 15;	Indels 2;	Gaps 1;

```
Oy 258 SGSPPKNASTPLGPKNISPPROVSVPRQSSPPKXVSPPPQA 302
      | | | | | | | | | | | | | | | | | | | |
Db 79 STSPPVATTPPALPKPLPP--LSPPQTTPPPPAITPPPPA 121
```

**RESULT 15**

Probable GT-11911 ethylhexyl DNA-binding protein [Imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: F96729  
 R:Theocystis, A., Exker, J.R., Palm, C.J., Federspiel, N.A., Kaul, S., White, O., Alonso,  
 Chin, C.R., Chung, M.K., Conn, L., Conway, A.B., Conway, A.R., Greasy, T.H., Dewar, K.,  
 Ansen, N.E., Hughes, B., Halzart, L.  
 Nature 406, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mateti, R.; Matzella,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Seizberg, S.V.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: AB6141; MUID:21016719  
 A:Accession: F96797  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-603 <STD>  
 A:Cross-references: GB:AE005173; NID:g11054438; PIDN:AA627801.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F7012.5  
 A:Map position: 1

Query Match	6.2%;	Score 107;	DB 2;	Length 603;
Best Local Similarity	17.08;	Pred. No. 1.9;		
Matches	50;	Conservative	54;	Mismatches 88;
				Indels 102;
				Gaps 11

```

Qy 97 RKSGCIVAVLEKLSPHNGSOAKVYRPHXSTMO-----OLESTINE-- 139
Db 97 KAKENGY---IRAKKCKEKENYATKYHRTYEGCTGCKSECKTYRFPDLALSOGT 132
Qy 140 -----PEPILADRPCKILKIV-----IRCHLALRLPISIDMESPL 179
Db 153 SLHHHQOTPLRPOQNNNNNNNNNNNSSTIRPYPTVMPLTSSIRPYQOINPWF 212
Qy 180 PULSEALSPATAKDETOMILSAARSELG-----MYSK 216
Db 213 PNTSDFLSDNSTS---SSSYSTSSDMGGGATITRRKKRRKKVFERPKLMQVYDK 268
Qy 217 KQEF---YLRRARR-----RKFAKPVLOSISEKMPVEFHTPMAYIRDSGSPKNAST 267
Db 269 QEQLRKFLAEVKEHEHRLVREESMR-----VQELARINEHEILAOERSMSAAKDAV 323
Qy 268 PSL-----GPPNISPPOVS-----VPORSSPPKNVSPPOP 301
Db 324 MAFLOKLEKOPNPQPOPOQVPSMLONNNOQPRQPSPPQPPAPVLPQ 377

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2002, 00:44:21 : Search time 24.18 Seconds  
(without alignments) 530.032 Million cell updates/sec

Title: US-09-828-313-35  
Sequence: 1 MGLPFCSCVTGQGVRRVYP.....SAASQGVQRNRGNKSLTYWA 331

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt 4.0

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	6.7	891	KIN1_SCHPO	P22987 schizosacch
2	116	6.7	912	PGCB_BOVIN	Q28062 bos taurus
3	112.5	6.5	569	STE7_SCHPO	Q10136 schizosacch
4	105	6.1	2845	APC_MOUSE	Q61315 mus muscu
5	103.5	6.0	830	DYN1_CAEEL	P39055 caenorhabd
6	103	5.9	1822	ZAP3_HUMAN	P49750 homo sapien
7	103	5.9	2843	APC_HUMAN	P25054 homo sapien
8	103	5.9	3703	ABF1_HUMAN	Q15911 homo sapien
9	102.5	5.9	864	WS14_MOUSE	Q99me3 mus muscu
10	102	5.9	620	EXTN_TOBAC	P13983 nicotiana t
11	101.5	5.9	1321	IRS2_MOUSE	P81122 mus muscu
12	101	5.8	1540	RM1_DROME	Q99p11 mus muscu
13	100	5.8	1943	PC15_MOUSE	P70478 ratu
14	100	5.8	2842	APC_MOUSE	P70478 ratu
15	100	5.8	3726	ABF1_MOUSE	Q08696 mus muscu
16	99.5	5.7	760	FXM1_MOUSE	Q9fpe6 chlamydomon
17	99	5.7	555	GPI_CHLRE	Q9fpe6 chlamydomon
18	99	5.7	759	FXM1_RAT	P97691 ratu
19	99	5.7	1021	YLB8_CAEEL	P46582 caenorhabd
20	99	5.7	3530	MY15_HUMAN	Q9ukr7 homo sapien
21	98.5	5.7	806	MK07_MOUSE	Q9fpe6 mus muscu
22	98	5.7	497	WAS2_HUMAN	Q9fpe6 mus muscu
23	98	5.7	678	T2D5_MOUSE	Q9fpe6 mus muscu
24	98	5.7	678	T2D5_MOUSE	Q9fpe6 mus muscu
25	97.5	5.6	1745	EX1L_MOUSE	Q9fpe6 mus muscu
26	97	5.6	306	EX1L_MOUSE	Q9fpe6 mus muscu
27	97	5.6	426	EX1L_TOBAC	Q9fpe6 mus muscu
28	97	5.6	630	EX1L_TOBAC	Q9fpe6 mus muscu
29	97	5.6	677	Y182_HUMAN	Q9fpe6 mus muscu
30	96.5	5.6	1157	Y182_HUMAN	Q9fpe6 mus muscu
31	96.5	5.6	1386	ZAP3_MOUSE	Q9fpe6 mus muscu
32	96	5.5	1487	BLM_DROME	Q9fpe6 mus muscu
33	96	5.5	1805	RM1_HUMAN	Q9fpe6 mus muscu

34	95.5	5.5	564	ATR_HUMAN	Q9h6x2 homo sapien
35	94.5	5.4	522	POLS_RUBIV	P08564 rubella vir
36	94.5	5.4	701	CG1_HUMAN	Q13495 homo sapien
37	94.5	5.4	732	POK_DROME	Q01842 drosophila
38	94.5	5.4	787	SMO_HUMAN	Q99835 homo sapien
39	94.5	5.4	815	MK07_HUMAN	Q13164 homo sapien
40	94	5.4	597	V70K_OTMV	P20130 ononis yell
41	94	5.4	1362	BRD4_HUMAN	Q60885 homo sapien
42	94	5.4	2715	TRX2_HUMAN	Q9um66 homo sapien
43	93.5	5.4	475	CAP1_HUMAN	Q01518 homo sapien
44	93	5.4	395	RRP2_DVU2	Q06427 newcastle d
45	93	5.4	504	YHV9_YEAST	P38854 saccharomyc

## ALIGNMENTS

RESULT	ID	Accession	Standard	PRT	891 AA.
1	KIN1_SCHPO	P22987	074392		
AC	KIN1_SCHPO	P22987	074392		
DT	01-AUG-1991 (Rel. 19, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DE	16-OCT-2001 (Rel. 40, Last annotation update)				
GN	Protein Kinase Kin1 (EC 2.7.1.-)				
GN	KIN1 OR SPK4F6.06				
OS	Schizosaccharomyces pombe (fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetes; Schizosaccharomycetaceae;				
OC	Schizosaccharomycetes.				
OX	NCBI_TaxID=4996;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91045979; PubMed=2236039;				
RA	Levin D.E., Bishop J.M.;				
RT	"A putative protein kinase gene (kin1+) is important for growth				
RT	polarity in Schizosaccharomyces pombe.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 87:8272-8276(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972;				
RA	Gwilliam R., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;				
RL	Submitted (Sep-1998) to the EMBL/GenBank/DBJ databases.				
CC	- FUNCTION: PROBABLE SERINE/THREONINE PROTEIN KINASE. IMPORTANT FOR				
CC	GROWTH POLARITY IN S.POMBE.				
CC	- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.				
CC	STRONG, TO YEAST KIN1 AND KIN2.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL: M64999; AAA63577.1; -				
DR	EMBL: AL031534; CA20726.1; -				
DR	PIR: A38903; A38903.				
DR	HSP: P00516; IPRK.				
DR	InterPro: IPR000719; Euk_pkinase.				
DR	InterPro: IPR001772; KAL.				
DR	InterPro: IPR002290; Ser_thr_pkinase.				
DR	InterPro: IPR001245; Tyr_pkinase.				
DR	Pfam: PF02149; KAL.1.				
DR	Pfam: PF00069; pkinase.1.				
DR	PRINTS: PR00109; TYRKINASE.				
DR	SMART: SM00220; S_TKc.1.				
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP.1.				
DR	PROSITE: PS00108; PROTEIN_KINASE_ST.1.				
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM.1.				
KW	Transferase; Serine/threonine-protein kinase; ATP-binding.				
FT	DOMAIN 125 395				
FT	PROTEIN KINASE.				

FT NP\_BIND 131 139 ATP (BY SIMILARITY).  
 FT BINDING 154 156 ATP (BY SIMILARITY).  
 FT ACT\_SITE 266 264 V->D (IN REF. 1).  
 FT CONFLICT 141 141 V->E (IN REF. 1).  
 FT CONFLICT 247 247 I->N (IN REF. 1).  
 FT CONFLICT 620 620 I->N (IN REF. 1).  
 FT CONFLICT 777 778 IR->ML (IN REF. 1).  
 SQ SEQUENCE 891 AA: 98748 MW: 8909BB6D82C0358 CRC64;

Query Match 6.7%; Score 117; DB 1; Length 891;  
 Best Local Similarity 21.8%; Pred. No. 0.3;

Matches 78; Conservative 50; Mismatches 137; Indels 92; Gaps 17;

QY 17 VYPDGIVEN--LSKSCSVHDLGPNFY-----VCGSTPYITNMAE----- 60  
 DB 262 VVHRDLKIENTILSKTGDIKIDFGISNLXRSQRTGFCGLFAAPBLNAPYIGPE 321  
 QY 61 -EVLEYGVYVCATPMAOPFLERQPKVH---RGSKILPFGSKGVHRELSPTHGS 115  
 DB 322 VVWMSFGIVLVLCGV--PPDQNMALHAKIKGTVEPSY-----LSSDCKGL 371  
 QY 116 QOSKRYED-YHSVTMOOLE---IRN-EGPEPHLAGRPSKHLKLVFIRHCLRLRLPR 169  
 DB 372 LSRMLVTDPLKRALLEVLNHPMIMIRNYEGPPASFPAPERSP----- 412  
 QY 170 ISIDMESPLNLGSGALSPATAKDEITOMILKSAARS--ELGMY-----VSKROEFL 222  
 DB 413 IITLPDPELTIEHMGPFQFQPEKIVRELKTVISSEAVGSLAKGFGYSGPNSADKKKSFE 472  
 QY 223 RRARRRRKFAKPYLVGISSEMKPYMEFHTPM-----AYRDSGSPKNAAP--S 269  
 DB 473 FRIHAADHDIENPLTLPISLNMNTDIYDAFHPLISIVYLSERVYEKGGMNRIAKTPVSS 532  
 QY 270 LRGPNISPPQVSVPPORSSPPKPNVSPPOPAFVARTASKYSAASQOVRNKGAK 326  
 DB 533 VP-----SSPYQ-----PTSYNRTLPMPREV---AYKGDSESPRVSRTSLAR 573

RESULT 2  
 PGCB\_BOVIN STANDARD; PRT: 912 AA.  
 AC 028062;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Brevican core protein precursor.  
 GN BCAN.  
 OS Bos taurus (bovine).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=94193597; PubMed=8144512;  
 RA Yamada H., Watanabe K., Shimonaka M., Yamaguchi Y.;  
 RT "Molecular cloning of brevican, a novel brain proteoglycan of the  
 RT aggrecan/versican family.";  
 RL J. Biol. Chem. 269:10119-10126(1994).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE TERMINALLY DIFFERENTIATING AND  
 CC THE ADULT NERVOUS SYSTEM DURING POSTNATAL DEVELOPMENT. COULD  
 CC STABILIZE INTERACTIONS BETWEEN HA AND BRAIN PROTEOGLYCANS.  
 CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY  
 CC SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: BRAIN; EXPRESSED IN CEREBELLAR ASTROCYTES BUT  
 CC NOT IN NEURONS.  
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X75887; CAA53481.1; -;  
 DR HSSP: P20693; 1HLJ.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR000742; EGF\_2.  
 DR Interpro: IPR003006; Iq\_Mc.  
 DR Interpro: IPR003596; Iq\_V.  
 DR Interpro: IPR00538; Link.  
 DR Interpro: IPR000436; Sushi\_SCR\_CCP.  
 DR Interpro: IPR001304; lectin\_C.  
 DR Pfam: PF00008; EGF\_1.  
 DR Pfam: PF00047; Iq; 1.  
 DR Pfam: PF00059; lectin\_C; 1.  
 DR Pfam: PF00084; sushi\_1.  
 DR Pfam: PF00193; Xlink; 2.  
 DR Pfam: PF00193; Xlink; 2.  
 DR ProDom: PD000918; Link; 2.  
 DR SMART: SM00032; CCP; 1.  
 DR SMART: SM00034; CLECT; 1.  
 DR SMART: SM00181; EGF; 1.  
 DR SMART: SM00406; IGV; 1.  
 DR SMART: SM00445; LINK; 2.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS00290; Iq\_Mc; 1.  
 DR PROSITE: PS01241; LINK; 2.  
 DR PROSITE: PS00615; C-TYPE-LECTIN\_1; 1.  
 DR PROSITE: PS50041; C-TYPE-LECTIN\_2; 1.  
 DR Glycoprotein: Hyaluronic acid; Proteoglycan; Lectin; Signal; Sushi;  
 KW EGF-like domain; Repeat; Immunoglobulin domain.  
 FT SIGNAL 1 22  
 FT CHAIN 23 912  
 FT DOMAIN 32 158  
 FT DOMAIN 174 251  
 FT DOMAIN 272 353  
 FT DOMAIN 647 683  
 FT DOMAIN 811 811  
 FT DOMAIN 812 876  
 FT DISULFID 57 137  
 FT DISULFID 179 250  
 FT DISULFID 203 224  
 FT DISULFID 277 352  
 FT DISULFID 301 322  
 FT DISULFID 651 662  
 FT DISULFID 656 671  
 FT DISULFID 673 682  
 FT DISULFID 689 700  
 FT DISULFID 717 809  
 FT DISULFID 785 801  
 FT DISULFID 816 859  
 FT DISULFID 845 872  
 FT DISULFID 130 150  
 FT CARBOHYD 337 357  
 SQ SEQUENCE 912 AA: 99554 MW: 677B3EB1C688C4D7 CRC64;

Query Match 6.7%; Score 116; DB 1; Length 912;  
 Best Local Similarity 24.2%; Pred. No. 0.37;  
 Matches 54; Conservative 26; Mismatches 101; Indels 42; Gaps 8;

QY 140 PEPHLAGDRSKHLKLVFIRHCLRLRLPR-----ISIDME-----SPLPMS 183  
 DB 368 PASHASDALE---AIVTVELTELKLPDQAVESRGAIVSPILEDGGSGSTPDP 424

QY 184 GEALSPATANKETITOMILKSAARSPLGMYVSKROEFLRRARRRK-----FA 232  
 DB 425 AEA--PRLLEFE--TOSIYPLPGSSSEBKGKLEOEDEKTRGEEKEEBEEDEADALMA 481  
 QY 233 WKPVLOSISEMKPMEFHPPMAYRDSGSPPKNA-----STPLSGPKNISPPROVSVPQ 286  
 DB 482 WPELSLSDPEALPTEPEVPEESLQASPPYRAALQPGVSPPYDEPAAPRPVILGPT 541  
 QY 287 RSSPPPKN---VSPPOPAFVARTASKYSAASQOVQNNRNNAK 326  
 DB 542 KTLPPREGNLASPPSTVLGARELEETGGEPLSGAPRESE 584

## RESULT 3

STEF\_SCHPO STANDARD; PRT: 569 AA.  
 ID STEF\_SCHPO Q10136;  
 AC 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE stef protein  
 GN ST7 OR SPAC23E2.03C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NX NCBI\_TaxID=4896;  
 RX MEDLINE=20296734; PubMed=10835379;  
 RA Matsuyama A., Yabana N., Matanabe Y., Yamamoto M.;  
 RT "Schizosaccharomycetes pombe Ste7p is required for both promotion and  
 RL withholding of the entry to meiosis.";  
 RT Genetics 155:539-549(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: HAS A ROLE IN PROMOTING MEIOSIS WHEREBY IT  
 CC IS INVOLVED IN ESTABLISHING THE MATING PHEROMONE SIGNALING  
 CC PATHWAY. IT ALSO HAS A ROLE IN SUPPRESSING MEIOSIS UNTIL THE  
 CC CONJUGATION PROCESS IS COMPLETE.  
 CC -1- INDUCTION: BY NITROGEN STARVATION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB036789; BAA90541.1; -  
 DR EMBL; Z68887; CA93115.3; -  
 KW Conjugation; Meiosis; Pheromone.  
 FT DOMAIN 200 261 SER-RICH.  
 FT DOMAIN 464 536 SER-RICH.  
 SO SEQUENCE 569 AA; 61101 MW; FDD23893E9A7A6AC CRC64;

Query Match 6.5%; Score 112.5; DB 1; Length 569;  
 Best Local Similarity 23.8%; Pred. No 0.37;  
 Matches 75; Conservative 35; Mismatches 120; Indels 85; Gaps 14;

QY 67 VTYFVCAITRAQPELEKPKVYHSGSILPRFSKHGV-HVRELSPTHSGQSKRVVDYH 125  
 DB 18 ITGKLEFETNA-----ATSPRSFQIOLIRINGVAVYHHGLSPSHKHPTASASDSD 69  
 QY 126 SVTWQOLESTRNCEPEPHLAGDRPSKH-----KLVIRIRCLR 163  
 DB 70 EAT-----EIFTHSQPLPIPAQPSKSKHAIDFKRFPKSKASSLPCSKSNDSMWICMLK 125

QY 164 ALRLPRIST-----DLMESPLPULSGEALSPR-----ATAK-----DEITOMILK 203  
 DB 126 AITSKRYALFGSSQAKAELIMVP--PMLACKPLSANSSFGSSAKTFQLPEDSVT-----P 180  
 QY 204 SAARSELGMYVSKROEFLRRARRRRRFAKPKVLOSISEMKPMEFHPPM-----AYR 256  
 DB 181 SAEIYQOPFSNPNAPITTSATHTSQF-----STSSSSSVNHTVWVNPVFFQYN 233  
 QY 257 DSGSPPKNASTPLSGPKNISPPROVSPQRSSPP--PNVSPPOPAFVARTASKYSA 314  
 DB 234 SSWTAPSSSSSSVAD-----FVPRROFSVSSASDPQTPISMSPPIPP-----FPSGSAR 284  
 QY 315 SOQVORNRGNKAKLY 329  
 DB 285 MYQNOQSYLSPOSHY 299

## RESULT 4

APC\_MOUSE STANDARD; PRT: 2845 AA.  
 ID APC\_MOUSE Q61315; Q62044;  
 AC 061315; Q62044;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Adenomatous polyposis coli protein (APC protein) (MAPC).  
 GN APC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RX MEDLINE=92263101; PubMed=1350108;  
 RA STRAIN=C57BL/6J; AND CAST/ET1; TISSUE=brain;  
 RC MEDLINE=92263101; PubMed=1350108;  
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,  
 RT Multiple intestinal neoplasia caused by a mutation in the murine  
 RT homolog of the APC gene.";  
 RT Science 256:668-670(1992).  
 RN [2]  
 RP ERRATUM.  
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,  
 RA Luongo C., Gould K.A., Dove W.F.;  
 RL Science 256:1114-1114(1992).  
 RN [3]  
 RP SEQUENCE OF 1-45 FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=Liver;  
 RA Dicker F., Lambert S., Reilmair A., Baillhausen W.G.;  
 RT "The murine APC gene: alternative splicing of 5' untranslated  
 RL region segments.";  
 RL Submitted (Oct-1993) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP ALTERNATIVE SPLICING.  
 RA MEDLINE=94061824; PubMed=8242607;  
 RX Oshima M., Sugiyama H., Kitagawa K., Taketo M.;  
 RT "APC gene messenger RNA: novel isoforms that lack exon 7.";  
 RL Cancer Res. 53:5589-5591(1993).  
 CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-  
 CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE  
 CC ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATEININ (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: FORMS HOMODIMERS AND ASSOCIATES WITH CATEININS (BY  
 CC SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,  
 CC LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.  
 CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M88127; AAB59632.1; -  
DR EMBL: U02937; AAA03443.1; -  
DR HSP: Q02248; 3BCT.  
DR MGD: MGI:88039; APC.  
DR InterPro: IPR000225; Armadillo.  
DR Pfam: PF00514; Armadillo\_seg. 6.  
DR SMART: SM00185; ARM. 5.  
DR PROSITE: PS50176; ARM\_REPEAT. 1.  
KW Anti-oncogene; Phosphorylation; Alternative splicing; Repeat;  
KW Coiled coil.  
FT DOMAIN 1 61 COILED COIL (POTENTIAL).  
FT DOMAIN 125 245 COILED COIL (POTENTIAL).  
FT DOMAIN 1 728 LEU-RICH.  
FT REPEAT 451 493 ARM 1.  
FT REPEAT 503 545 ARM 2.  
FT REPEAT 546 589 ARM 3.  
FT REPEAT 590 636 ARM 4.  
FT REPEAT 637 681 ARM 5.  
FT REPEAT 682 723 ARM 6.  
FT REPEAT 724 765 ARM 7.  
FT DOMAIN 739 2834 SER-RICH.  
FT DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 1664 1891 HIGHLY CHARGED.  
FT VARSPLIC 243 276 MISSING (IN ISOFORM 3 AND ISOFORM 4).  
FT VARSPLIC 310 410 MISSING (IN ISOFORM 3 AND ISOFORM 4).  
FT VARSPLIC 120 120 V -> A (IN STRAIN CAST/EI).  
FT VARSPLIC 493 493 V -> I (IN STRAIN CAST/EI).  
FT VARSPLIC 797 797 Y -> F (IN STRAIN CAST/EI).  
FT VARSPLIC 1330 1330 A -> T (IN STRAIN CAST/EI).  
FT VARSPLIC 1618 1618 A -> S (IN STRAIN CAST/EI).  
FT VARSPLIC 2294 2294 G -> A (IN STRAIN CAST/EI).  
FT VARSPLIC 2496 2496 H -> Q (IN STRAIN CAST/EI).  
FT VARSPLIC 2523 2523 T -> A (IN STRAIN CAST/EI).  
FT VARSPLIC 2813 2813 T -> S (IN STRAIN CAST/EI).  
SQ SEQUENCE 2843 AA; 311086 MW; 145CA73CFF570A499 CRC64;

Query Match 6.1%; Score 105; DB 1; Length 2845;

Best Local Similarity 21.7%; Pred. No. 10;  
Matches 56; Conservative 42; Mismatches 110; Indels 50; Gaps 9;

QY 79 PFLERQPKVHHGSKLLPRFSKHGVHRELRSPTHSOOSRKVEDYSHVMQOLES---I 135  
DB 2152 PGOEKPFTSNKGPRLKPGCKSTLEKKKIESNCKIKGCKV--YSLITGKIRNSSEI 2209  
QY 136 RNDEPPEHLADGRPSKHLKLVFIRHCLRLRLPRISIDLESPLNT--SGEALSPATA 193  
DB 2210 SSQMKQP-----LPTNMPISIRGRTMHIIFGLNSSSSSTSPVSKK 2249  
QY 194 KDIITMILKSAARSELGVYKQEFYLRARRRRKFWKVLQSTSEKMPVMEFTTP 253  
DB 2250 GPPKLPKAPKSSSEEP-GATTSPP-----GTRKA--GKSELSPITRQTSQI 2292  
QY 254 AYDSSGSPPKNA--STPSLPGPKNISPPQVSVQSSPPPKNNSPPPKAPFAVARTASKY 311  
DB 2293 SGNKSGSSSGSRDSTPSRPPQPLSRPKQSPGRNISPGRNISPKNLSQLPRITSSPS 2352  
QY 312 SAASQOVORNRGNMAKSLY 329  
DB 2353 YAST-----KSSSGSKMSY 2366

RESULT 5  
ID DYNL CAEEL STANDARD; PRT; 830 AA.  
AC P39055;  
DT 01-FEB-1995 (rel. 31, Created)

DR 30-MAY-2000 (rel. 39, Last sequence update)  
DE 16-OCT-2001 (rel. 40, Last annotation update)  
DT DYN-1.  
CN

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;  
OC Rhabditidae; Pelodetinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2.  
RX MEDLINE=97439883; PubMed=9294229;  
RA Clark S.G., Shurland D.L., Meyerowitz E.M., Bargmann C.I.,  
RA van der Bliek A.M.;  
RT "A dynamin GTPase mutation causes a rapid and reversible temperature-  
RT inducible locomotion defect in C. elegans";  
RT Proc. Natl. Acad. Sci. U.S.A. 94:10438-10443(1997).  
RN [2]  
RP REVISIONS TO C-TERMINUS.  
RC STRAIN-BRISTOL N2;  
RA van der Bliek A.M.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DDI databases.  
CC -1- FUNCTION: MICROTUBULE-ASSOCIATED FORCE-PRODUCING PROTEIN INVOLVED  
CC IN PRODUCING MICROTUBULE BUNDLES AND ABLE TO BIND AND HYDROLYZE  
CC GTP. MOST PROBABLY INVOLVED IN VESICULAR TRAFFICKING PROCESSES, IN  
CC PARTICULAR ENDOCYTOSIS.  
CC -1- CATALYTIC ACTIVITY: GTP + H(2)O = GDP + phosphate.  
CC -1- SUBCELLULAR LOCATION: MICROTUBULE-ASSOCIATED.  
CC -1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L29031; AAB72228.2; -  
DR HSP: Q05193; 2DYN  
DR InterPro: IPR001401; Dynamlin.  
DR InterPro: IPR000375; Dynamlin\_Central.  
DR InterPro: IPR003130; GED.  
DR InterPro: IPR001849; PH.  
DR Pfam: PF00350; dynamlin.1.  
DR Pfam: PF01031; dynamlin.2; 1.  
DR Pfam: PF02212; GED. 1.  
DR Pfam: PF00169; PH. 1.  
DR PRINTS: PR00195; DYNAMIN.  
DR SMART: SM00053; DYNCL. 1.  
DR SMART: SM00302; GED. 1.  
DR SMART: SM00233; PH. 1.  
DR PROSITE: PS00410; DYNAMIN. 1.  
DR PROSITE: PS50003; PH\_DOMAIN. 1.  
KW Hydrolyase; Motor protein; GTP-binding; Microtubules; Multigene family;  
KW Endocytosis.  
FT NP\_BIND 40 47 GTP (BY SIMILARITY).  
FT NP\_BIND 138 142 GTP (BY SIMILARITY).  
FT NP\_BIND 207 210 GTP (BY SIMILARITY).  
FT DOMAIN 519 624 PH.  
SQ SEQUENCE 830 AA; 93348 MW; FC3D7106D079EDC5 CRC64;

Query Match 6.0%; Score 103.5; DB 1; Length 830;

Best Local Similarity 24.2%; Pred. No. 2.8;  
Matches 52; Conservative 32; Mismatches 84; Indels 47; Gaps 10;

QY 98 FSKHGVHRELRSPTHSOOSRKVEDYSHVMQOLESIRNDEPPEHLADGRPSKHLKLVF 157  
DB 619 FLRAGVPEPKQAEDESDQEMEDSIDPQLRQVETIRN-----LVDSY 663  
QY 158 INHCLRLRL--LPRISIDLESPLNLSGEALSPATATAKDEITQMLK-----SAARSELG 211



```

Db      664 MRLTKTKDLVPRVAVHNL-----IVNOTGEFM-----KDLHLHLQCCDITLALMEESQ 713
QY      212 MYVSKROEFYLRARRRRKFAKPVQLQSTF--MKPVMEFTPMAYRDSGSSPPKNASTPS 269
Db      714 IENQREEML-----RMVHACKREALPIISEVMSTLDDQPPPLPMSDYRPHSGSPVP 767
QY      270 LPPKNISSP--ROVSVPOR---SSPPKNSVSP 299
Db      768 RPAF---APGGRQAPMPRPGGAPPPGCMRPP 799

```

```

RESULT 6
ZAP3_HUMAN
ID ZAP3_HUMAN STANDARD; PRT; 1822 AA.
AC P49750; P49752; Q9P1V7;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear protein ZAP3 (ZAP113).
GN ZAP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Baradaran L., Birditt B.,
RA Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.;
RT "Sequencing of human chromosome 14q24.3 region."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
RC TISSUE=Brain;
MEDLINE=95319502; PubMed=7596406;
RA Sherrington R., Rogaev E.I., Liang Y., Rogaeva E.A., Levesque G.,
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
RA Fonci J.-F., Bruni A.C., Montesi M.P., Sobri S., Rainero I.,
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA Sansone P., Polinsky R.U., Masco W., da Silva H.A.R., Haines J.L.,
RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA Rommens J.M., St. George-Hyslop P.H.;
RT "Cloning of a gene bearing missense mutations in early-onset familial
RT Alzheimer's disease."
RL Nature 375:754-760(1995).
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1661.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AC007956; AAF61275.1;
DR EMBL: L40403; AAC42008.1; ALT_FRAME.
DR EMBL: L40403; AAC42006.1;
KV Nuclear protein.
FT DOMAIN 15 205 PRO-RICH.
FT DOMAIN 382 430 GLN-RICH.
FT DOMAIN 807 1209 ARG-RICH.
FT DOMAIN 1488 1577 ARG-RICH.
FT CONFLICT 621 621 P -> S (IN REF. 2).
FT CONFLICT 1404 1404 T -> I (IN REF. 2).
FT CONFLICT 1821 1821 K -> E (IN REF. 2).
SQ SEQUENCE 1822 AA; 204947 MW; 8E6CBB3FF5540C7D2 CRC64;

```

Query Match

5.9%; Score 103; DB 1; Length 1822;

```

Best Local Similarity 25.1%; Pred. No. 8.1;
Matches 47; Conservative 30; Mismatches 76; Indels 34; Gaps 9;
QY 159 RHCRLRLP--RISIDLMESPLPNTLSEALSPATAKDEITOMILKSAARSELGMVSK 216
Db 1217 RHPMAMHMSHSHSEEMMGASLSDQGLGYMTVLSORHETILKAA--QELKMLREQ 1274
QY 217 ROEFYLRARRRRKFAKPV-----LQISEMKPVMEFTPMAYR---DSGSP 262
Db 1275 KEQL-----QKMDFGSEFQMDHLPPQESRLQNTSS-RPGM-YPPPGSRPPPMCKVP 1327
QY 263 KNASTPLPGPKNISPPROVSVQORSPPKNSVSPPOPAFYARATSKVSAASQOYQNR 322
Db 1328 GSIVRPSAPPARSSVPTVTPVPIPPPPPPPL-PPPPVIRQISA-----VQGR 1378
QY 323 GNASKLY 329
Db 1379 WDEDSFY 1385

```

```

RESULT 7
APC_HUMAN
ID APC_HUMAN STANDARD; PRT; 2843 AA.
AC P25054; Q15162; Q15163;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein).
GN APC OR DP2.5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,
RA Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McCreich D.,
RA Flinnleat R., Markham A., Groffen J., Boguski M.S., Altschul S.F.,
RA Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.;
RT "Identification of FAP locus genes from chromosome 5q21."
RL Science 253:661-665(1991).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal Brain;
MEDLINE=91335210; PubMed=1678319;
RA Joslyn G., Carlson M., Thilveris A., Albertsen H., Gelbert L.,
RA Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,
RA Sargeant L., Krapcho K., Wolff E., Burt R., Hughes J.P.,
RA Warrington J., McPherson J.D., Wasmuth J., Le Fautier D.,
RA Aderrahim H., Chen D., Leppert M., White R.;
RT "Identification of deletion mutations and three new genes at the
RT familial polyposis locus."
RL Cell 66:601-613(1991).
RN [3]
RP ASSOCIATION WITH CATENINS.
RA MEDLINE=94082295; PubMed=8259519;
RA Su L.-K., Vogelstein B., Kinzler K.W.;
RT "Association of the APC tumor suppressor protein with catenins."
RL Science 262:1734-1737(1993).
RN [4]
RP REVIEW ON VARIANTS.
RA MEDLINE=94154728; PubMed=8111410;
RA Nagase H., Nakamura Y.;
RT "Mutations of the APC (adenomatous polyposis coli) gene."
RL Hum. Mutat. 2:425-434(1993).
RN [5]
RP VARIANTS FAP.
RA MEDLINE=91335211; PubMed=1651563;
RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
RA Petersen G., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;

```

RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer  
 RT patients.";  
 RT Science 253:665-669(1991).  
 RL [6]  
 RN VARIANTS FAP.  
 RX MEDLINE-93265030; PubMed-1338904;  
 RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,  
 RA Miki Y., Mori T., Nakamura Y.;  
 RT "Somatic mutations of the APC gene in colorectal tumors: mutation  
 RT cluster region in the APC gene";  
 RT Hum. Mol. Genet. 1:229-233(1992).  
 RL [17]  
 RN VARIANTS FAP.  
 RX MEDLINE-93244793; PubMed-1338691;  
 RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,  
 RA Nakamura Y., Horii A.;  
 RT "Somatic mutation of the APC gene in gastric cancer: frequent  
 RT mutations in very well differentiated adenocarcinoma and signet-ring  
 RT cell carcinoma";  
 RT Hum. Mol. Genet. 1:559-563(1992).  
 RL [18]  
 RN VARIANT FAP W-1348, AND VARIANTS D-1118; M-1292; V-1304 AND S-2502.  
 RX MEDLINE-93250848; PubMed-1338764;  
 RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,  
 RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,  
 RA Baba S., Nakamura Y.;  
 RT "Screening for germ-line mutations in familial adenomatous polyposis  
 RT patients: 61 new patients and a summary of 150 unrelated patients";  
 RT Hum. Mutat. 1:467-473(1992).  
 RL [19]  
 RN VARIANT FAP TRP-99.  
 RX TISSUE-Peripheral blood lymphocytes;  
 RC MEDLINE-95134544; PubMed-783149;  
 RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Ammann T., Roth J.,  
 RA Mueller A., Meiller H., Scott R.J.;  
 RT "Mutational analysis of the first 14 exons of the adenomatous  
 RT polyposis coli (APC) gene";  
 RT Eur. J. Cancer 30A:1709-1713(1994).  
 RL [10]  
 RN VARIANT FAP GLY-722.  
 RX MEDLINE-95135430; PubMed-7839331;  
 RA Stella A., Montero M., Reata N., Marchese C., Susca F., Gentile M.,  
 RA Romo L., Pilla S., Prete F., Martini C., Guanti G.;  
 RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in  
 RT FAP patients";  
 RT Hum. Mol. Genet. 3:1687-1688(1994).  
 RL [11]  
 RN ERRATUM.  
 RA Stella A., Montero M., Reata N., Marchese C., Susca F., Gentile M.,  
 RA Romo L., Pilla S., Prete F., Martini C., Guanti G.;  
 RT Hum. Mol. Genet. 3:1918-1918(1994).  
 RL [12]  
 RN VARIANT FAP ILE-171.  
 RX MEDLINE-97144176; PubMed-8990002;  
 RA van der Luijt R.B., Meera Khan P., Vaasen H.F.A., Tops C.M.J.,  
 RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,  
 RA Plug R.J., Griffioen G., Fodde R.;  
 RT "Molecular analysis of the APC gene in 105 Dutch kindreds with  
 RT familial adenomatous polyposis: 67 germline mutations identified by  
 RT DGGE, PTM, and southern analysis";  
 RT Hum. Mutat. 9:7-16(1997).  
 RL [13]  
 RN VARIANT LYS-1307.  
 RX MEDLINE-98400248; PubMed-9731522;  
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,  
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,  
 RA Antin-Ozertoks D., Andrulis I., Daly M., Pinsky L., Schrag D.,  
 RA Gallinger S., Kaback W., King M.C., Woodage T., Brody L.C.,  
 RA Gollwin A., Warner E., Weber B., Foulkes W., Offit K.;  
 RT "The APC I1307K allele and breast cancer risk";  
 RT Nat. Genet. 20:13-14(1998).  
 RL [14]  
 RN VARIANTS LYS-1307 AND GLN-1317.

RC TISSUE-Peripheral blood;  
 RX MEDLINE-98393712; PubMed-9724771;  
 RA Freyling I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,  
 RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,  
 RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;  
 RT "The APC variants I1307K and E1317Q are associated with colorectal  
 RT tumors, but not always with a family history";  
 RT Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).  
 RL [15]  
 RN VARIANT LYS-1307.  
 RX MEDLINE-98400259; PubMed-9731533;  
 RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,  
 RA Macdams M., Laken S.J., Tucker M.A., Brody L.C.;  
 RT "The APC I1307K allele and cancer risk in a community-based study of  
 RT Ashkenazi Jews";  
 RT Nat. Genet. 20:62-65(1998).  
 RL [16]  
 RN VARIANT LYS-1307.  
 RX MEDLINE-99138651; PubMed-9973276;  
 RA Gryfe R., Di Nicola N., Lal G., Gallinger S., Redston M.;  
 RT "Inherited colorectal polyposis and cancer risk of the APC I1307K  
 RT polymorphism";  
 RT Am. J. Hum. Genet. 64:378-384(1999).  
 RL [17]  
 RN VARIANTS FAP CYS-1171 AND THR-2738, AND VARIANTS GLY-1057 AND  
 RP VAL-1822.  
 RX MEDLINE-99133859; PubMed-9950360;  
 RA Wallis Y.L., Morton D.G., McKeown C.M., Macdonald F.;  
 RT "Molecular analysis of the APC gene in 205 families: extended  
 RT genotype-phenotype correlations in FAP and evidence for the role of  
 RT APC amino acid changes in colorectal cancer predisposition";  
 RT J. Med. Genet. 36:14-20(1999).  
 RL [18]  
 RN VARIANT FAP PRO-1184.  
 RX MEDLINE-99401091; PubMed-10470088;  
 RA Lamium H., Ilyas M., Rowan A., Clark S., Johnson V., Bell J.A.,  
 RA Freyling I.M., Efstathiou J., Pack K., Payne S., Roylance R.,  
 RA Gorman P., Sheer D., Neale K., Phillips R., Talbot I.C., Bodmer W.F.,  
 RA Tomlinson I.P.M.;  
 RT "The type of somatic mutation at APC in familial adenomatous polyposis  
 RT is determined by the site of the germline mutation: a new facet to  
 RT Knudson's 'two-hit' hypothesis";  
 RT Nat. Med. 5:1071-1075(1999).  
 RL [19]  
 RN -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-  
 CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE  
 CC  
 CC -1- SUBUNIT: FORMS HOMOLOGOMERS AND ASSOCIATES WITH CATENIN.  
 CC  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES TYPES.  
 CC  
 CC -1- PTM: PHOSPHORYLATED BY GSK-3B.  
 CC  
 CC -1- DISEASE: DEFECTS IN APC ARE A CAUSE OF FAMILIAL ADENOMATOUS  
 CC POLYPOSIS (FAP) AND GARDNER'S SYNDROME (GS), THAT CONTRIBUTE TO  
 CC TUMOR DEVELOPMENT IN PATIENTS WITH NONINHERITED FORMS OF  
 CC COLORECTAL CANCER. FAP IS CHARACTERIZED BY ADENOMATOUS POLYPS OF  
 CC THE COLON AND RECTUM, BUT ALSO OF UPPER GASTROINTESTINAL TRACT  
 CC (AMPULLARY, DUODENAL, AND GASTRIC ADENOMAS). THIS IS A VICIOUSLY  
 CC PREVALANT DISEASE WITH ONE OR MORE POLYPS PROGRESSING THROUGH  
 CC DYSPLASIA TO MALIGNANCY IN UNTREATED GENE CARRIERS WITH A MEDIAN  
 CC AGE AT DIAGNOSIS OF 40 YEARS.  
 CC  
 CC -1- DISEASE: APC MUTATIONS HAVE LED TO SOME INTERESTING OBSERVATIONS.  
 CC (1) THE GREAT MAJORITY OF THE MUTATIONS FOUND TO DATE WOULD RESULT  
 CC IN TRUNCATION OF THE APC PRODUCT. (2) ALMOST ALL THE MUTATIONS  
 CC HAVE OCCURRED WITHIN THE FIRST HALF OF THE CODING SEQUENCE, AND  
 CC SOMATIC MUTATIONS IN COLORECTAL TUMORS ARE FURTHER CLUSTERED IN A  
 CC PARTICULAR REGION, CALLED MCR (MUTATION CLUSTER REGION). (3) MOST  
 CC IDENTIFIED POINT MUTATIONS IN THE APC GENE ARE TRANSITIONS FROM  
 CC CYTOSINE TO OTHER NUCLEOTIDES. (4) THE LOCATION OF GERM-LINE  
 CC MUTATIONS TENDS TO CORRELATE WITH THE NUMBER OF COLORECTAL POLYPS  
 CC IN FAP PATIENTS. INACTIVATION OF BOTH ALLELES OF THE APC GENE  
 CC SEEMS TO BE REQUIRED AS AN EARLY EVENT TO DEVELOP MOST ADENOMAS  
 CC AND CARCINOMAS IN THE COLON AND RECTUM AS WELL AS SOME OF THOSE IN  
 CC THE STOMACH.











```
OY 79 PLELPKPVHNGSKILPRFSKHGVHRELRSPTGSGQSKRYEDHYSVTMOLES---I 135
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2152 PDGEKPFHSGPILKPGKSTLEAKKIESENKIGKRV--YKSLTIGKINSSEI 2209
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 136 RNCGRPHLAGRPSKHLKLVFIRHCLRALRLPRT-----SIDLMESPL--PILSGE 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2210 SSGMKOP-LQTNMS-----ISRGRTMHLPGVGRNSSSTFVSKKPKLTPAKNSP 2261
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 186 ALSPTVANDETIQMLKSAARSELGMYSKROEFLRRARRRRKFAKPVLOJISEMKP 245
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2262 SEGVVATTSPPRT-----KPAVKSLE-----SEPTROTSHI-- 2292
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 246 VMEFHPHMYARVSGS---PPKNA---STPLSGPKNISPPQVSVPORSSPPKKNVSP 299
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2293 -----SSGNKGPSSKSGSRDSTPSRPTQPLSRPMQSGRMSISPGKNGISTPN 2340
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 300 QPAFVARTAKSKSAASQVQVRNGKNKSLX 329
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2341 KLSQLPRITSSPSTAST-----KSSGSGKMSI 2366
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 15
ABFL_MOUSE STANDARD; PRT: 3726 AA.
ID ABFL_MOUSE
AC 061329;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-fetoprotein enhancer binding protein (Af motif-binding factor)
(At-binding transcription factor 1).
GN ABFL.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/MK X ICR: TISSUE=Brain;
RX MEDLINE=96194902; PubMed=8654949;
RA Ito A., Miura Y., Watanabe M., Sakai M., Inoue Y., Miki T.,
RA Hashimoto Y., Morinaga T., Nishi S., Yamaki I.;
RT "Cloning of the cDNA encoding the mouse ABFL1 transcription factor.";
RT Gene 168:227-231(1996).
CC -1- FUNCTION: Transcriptional activator that binds to the AT-rich core
CC sequence of the enhancer element of the AFP gene.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 4 HOMEBOX DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D26046; BAA05046.1; -.
DR HSP: P06601; IFTL.
DR MGD: MGI:99948; Abfl.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR003604; Znf_U1.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00046; homeobox; 4.
DR Pfam: PF00096; zf-C2H2; 20.
DR SMART: SM00389; HOX; 4.
DR SMART: SM00355; znf_C2H2; 22.
DR SMART: SM00451; znf_U1; 7.
DR PROSITE: PS00027; HOMEBOX 1; 2.
DR PROSITE: PS00071; HOMEBOX 2; 4.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 15.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 9.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
```

```
KW DNA-binding; Homeobox; Nuclear protein; Repeat.
FT ZN_FING 79 103 C2H2-TYPE.
FT ZN_FING 282 305 C2H2-TYPE.
FT ZN_FING 641 664 C2H2-TYPE.
FT ZN_FING 672 695 C2H2-TYPE.
FT ZN_FING 727 751 C2H2-TYPE.
FT ZN_FING 805 829 C2H2-TYPE (ATYPICAL).
FT ZN_FING 946 969 C2H2-TYPE (DEGENERATE).
FT ZN_FING 985 1009 C2H2-TYPE (ATYPICAL).
FT ZN_FING 1041 1065 C2H2-TYPE (ATYPICAL).
FT ZN_FING 1089 1113 C2H2-TYPE (ATYPICAL).
FT ZN_FING 1233 1256 C2H2-TYPE (ATYPICAL).
FT ZN_FING 1262 1285 C2H2-TYPE (ATYPICAL).
FT ZN_FING 1370 1395 C2H2-TYPE.
FT ZN_FING 1411 1433 C2H2-TYPE.
FT ZN_FING 1439 1462 C2H2-TYPE.
FT ZN_FING 1555 1579 C2H2-TYPE.
FT ZN_FING 1606 1630 C2H2-TYPE.
FT ZN_FING 1990 2013 C2H2-TYPE.
FT ZN_FING 2192 2211 HOMEBOX 1.
FT DNA_BIND 2192 2211 HOMEBOX 2.
FT DNA_BIND 2249 2308 C2H2-TYPE (ATYPICAL).
FT ZN_FING 2335 2358 C2H2-TYPE.
FT ZN_FING 2539 2561 HOMEBOX 3.
FT DNA_BIND 2650 2709 HOMEBOX 4.
FT ZN_FING 2720 2743 C2H2-TYPE.
FT DNA_BIND 2952 3011 HOMEBOX 4.
FT ZN_FING 3032 3056 C2H2-TYPE.
FT ZN_FING 3552 3576 C2H2-TYPE.
FT ZN_FING 461 491 POLY-ALA.
FT ZN_FING 771 785 POLY-ALA.
FT DOMAIN 1314 1317 POLY-ALA.
FT DOMAIN 1734 1748 POLY-GLN.
FT DOMAIN 1794 1799 POLY-GLN.
FT DOMAIN 1856 1863 POLY-GLN.
FT DOMAIN 2044 2059 POLY-PRO.
FT DOMAIN 2405 2408 POLY-PRO.
FT DOMAIN 3216 3220 POLY-ALA.
FT DOMAIN 3380 3409 POLY-PRO.
FT DOMAIN 3412 3420 POLY-GLN.
FT DOMAIN 3534 3550 POLY-GLN.
FT DOMAIN 3620 3623 POLY-PRO.
FT DOMAIN 3659 3662 POLY-SER.
SQ SEQUENCE 3726 AA; 406567 MW; 915ACBE588A72C98 CRC64;

Query Match 5.8%; Score 100; DB 1; Length 3726;
Best Local Similarity 27.5%; Pred. No. 33;
Matches 41; Conservative 23; Mismatches 59; Indels 26; Gaps 7;

OY 171 SIDLM-----SPLPMLSGEALSPATKDEITOMILK-SAARSELGMYVSKROE 219
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2375 SMDAMELILPTSSSGSTPMPSQATSTPAPSAAMANTAPSAFLQJLATDELTAFNSKAE- 2433
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 220 FYLRARRRRRRKFAKPVLIQ-----SISEKKPYMEFHHPMAVRDSSGP-PK--MASTPRL 270
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2434 ----ASDEKPKQADPPSAQPNQTEKGQKQPKPKQQLQLEQKTNAPQKRLPQPAAPSL 2489
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 271 PGPKNISPPQVSVPORSSPPKKNVSP 299
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2490 POPPPQAPPPQCPQLPQ--SSPSPQLSHLP 2517
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: September 16, 2002, 00:51:59
Job time: 458 sec
```





```

Db 101 NVGSPVHVNNHNAVSSMPSPDMDPADRRPOEPQARRPLSVATDNNMLEFFKDKDGMGVAMD 160
Qy 110 S---PTNGSOQSKRYVDYHSVTMOOLESRNEGPEPLADRRSKLKLVTIRHCLRLRLR 166
Db 161 TNWVARRGSSAGRKV-----SCAPSPQAPPAELAPLPS-----198
Qy 167 LPRISIDLMESPPLNLSGEALSP---TATAKDEITOMILKSAAR-----207
Db 199 LPEQGLDSPAAPALSPSGGLQIQGPERTSTTKSK--ELSPGSAQKSPSSOGTACACTQ 256
Qy 208 -----SFLGMVYSKROEYFLRRARRRRKPFAMKPVLOSISEMKPVMEFHTP--MAYRDSG 259
Db 257 PEAQPGAPGAPGASPSPOPPADOSPHTLRK-----VSKKLAPIPKVPFGQPGAMADQASG 311
Qy 260 SPPK---NASTPSLPGFKNISPPROVSV-----GRSSPPK- 293
Db 312 QPSVSLSPPTSPSPSYGLSYFGQYSLASQGLSPAAPLAPLSPSVFTSLKSRPTKRP 371
Qy 294 ---NVSPPOPAFAVARTAS 309
Db 372 RORPTLPPOPPPTVNLAS 390

```

```

RESULT 2
ID 075160 PRELIMINARY; PRT; 818 AA.
AC 075160;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE KIA0672 PROTEIN.
GN KIA0672.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL: AB014572; BAA31647.1; .
DR HSSP: Q07960; IJGP.
DR InterPro: IPR000198; RhoGAP.
DR Pfam: PF00620; RhoGAP; 1.
DR SMART: SM00324; RhoGAP; 1.
SQ SEQUENCE 818 AA; 89262 MW; 996CBCE9B523808 CRC64;

```

```

Query Match 7.3% Score 126; DB 4: Length 818;
Best Local Similarity 20.7%; Pred No. 0.012;
Matches 80; Conservative 55; Mismatches 140; Indels 112; Gaps 16;

```

```

Qy 3 LTFPSCVTVQGVYRVYP--DGHVENLSKSCSVH-----DILLGNPYVCGSTPYTTIN 55
Db 397 MFPSNMAIYLG-PNLMLPQAEQNTTEMTVLSIQIYGLIEPILOHADWFPGETIEFNIG 455
Qy 56 RMAAEVLEVYGVYVCACTPAOPEFLEROKKYVNR-----GSKILRFSGHGVHREL 109
Db 456 NIGSPYVHNNANNSMSPDMDPADRRPOEPQARRPLSVATDNNMLEFFKDKD--LRKIQ 513
Qy 110 S-----PTNGSOQSKRYVDYHSVTMOOLESRNEGPEPLADRRSKLKLVTI 158
Db 514 SMGVYVMDTNWVARRGSSAGRKV-----SCAPSPQAPPAELAPLPS-----559
Qy 159 RHCLRLRLPRISIDLMESPPLNLSGEALSP---TATAKDEITOMILKSAAR-----207
Db 560 -----LPEQGLDSPAAPALSPSGGLQIQGPERTSTTKSK--ELSPGSAQKSPSSO 609

```

```

Qy 208 -----SELGMVYSKROEYFLRRARRRRKPFAMKPVLOSISEMKPVMEFHTP 253
Db 610 GTACACTQGAQPGAPGASPSPOPPADOSPHTLRK-----VSKKLAPIPKVPFGQPG 664
Qy 254 AYRDSG---SPPKNAST-PSLPGFKNISPPROVSV-----Q 286
Db 665 AMADQASQGLSPSVSLSPPTSPSYGLSYFGQYSLASQGLSPAAPLAPLSPSVFTSTLS 724
Qy 287 RSSPPK---NVSPPOPAFAVARTAS 309
Db 725 KSRPTKPPORPTLPPOPPPTVNLAS 751

```

```

RESULT 3
ID 09VZ78 PRELIMINARY; PRT; 1381 AA.
AC 09VZ78;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE CG14964 PROTEIN.
GN CG14964.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Cealicher S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.W., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Jiaoli B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Laake P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Pui V., Reese M.G.,
RA Palazzolo M., Peltman G.S., Pan S., Saunders R.D.C., Scheeler F., Shen H.,
RA Reinelt K., Remington K., Sanders R.D.C., Skupski M., Smith T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003477; AAF47729.1; .
DR HSSP: P56276; ITLK.
DR FlyBase: FBgn0035410; CG14964.

```

DR InterPro: IPR003962; FNIII\_repeat.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003599; Ig.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR Pfam: PF00047; Ig; 2.  
 DR Pfam: PF00047; Ig; 2.  
 DR PRINTS: PR00014; FNTYPELIT.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 DR SMART: SM00060; FN3; 2.  
 DR SMART: SM00409; IG; 2.  
 DR Repeat:  
 KW  
 SO SEQUENCE 1381 AA; 135482 MW; 8341F17BB4542745 CRC64;

Query Match 7.2%; Score 125.5; DB 5; Length 1381;  
 Best Local Similarity 22.4%; Pred. No. 0.025;  
 Matches 72; Conservative 46; Mismatches 135; Indels 69; Gaps 15;

QY 30 SCVHOLLGNP-DYVCGSTPYTITRNMAAEVL-----ETGVYFCATPPMAQPLELR 83  
 DB 367 STIHLDLIESEYKRVKAENPYGLSEPSSELELTPPKRGITRKPSATRIAGDEKDK 426  
 QY 84 QPKVYHSGSKILPRESKGVHVELRSPTHGSOOSRKVFYDHSVTMOOLESTRNGEPH 143  
 DB 427 -PKTGAGGMQVPPR-----RKTLSPPRQADAST-----GMSPKQSPSAKRK-PKQ 471  
 QY 144 LAGDRPSKH-ILKLVFIRHCLNALRLPRISIDLMESPLPMLSGEALSPATAKDEITQML 202  
 DB 472 LIDNEQLTHMSYGSIDALK-----MDVKRSPSLMSAUSAKPPTDSSNKLMLTL 523  
 QY 203 KSAARSELCGVYSKROEYFLRRARRRRFAMKPVLOSISEMVMVEFTPMAYRDSGP- 261  
 DB 524 VTTTLAPLRKSVP-----SPKAPRTPATSPKLKFN-----KPSGAP 561  
 QY 262 -----PKNSTPSLPGPKNIS--PPROVSVPOSSPPKNSP-PPQAFPA-----R 306  
 DB 562 DRSPVQPKQPLRPTPMETPKKASPNKRSLSPPKRQPPPLKRSPTPEPIKVYALLR 621  
 QY 307 TASKYS-AAQQOVQNRGNKAS 327  
 DB 622 SAEPVQLGVNQVRRRFSQTL 643

RESULT 4  
 Q9LQ09 PRELIMINARY: PRT: 796 AA.  
 AC 09LQ09:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE F16P17.12 PROTEIN.  
 GN F16P17.12.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxId=3702;  
 QX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Sakano H., Liu S.X., Yu G., Lee J.M., Lenz C., Pham P., Toriumi M.,  
 RA Chin C., Chlou J., Choi E., Chung J.M., Gonzalez A., Homing B., Liu A.,  
 RA Vaynsberg M., Altieri H., Brooks S., Buehler E., Chao Q., Conn L.,  
 RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,  
 RA Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,  
 RA Becker J.R., Federspiel N.A., Theologis A.;  
 RT "The sequence of BAC F16P17 from Arabidopsis thaliana chromosome 1.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC011000; AAF75808.1;  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR Pfam: PF00226; DnaJ\_1.

DR PRINTS: PR01217; PRICHEXTENS.  
 DR SMART: SM00271; DnaJ\_1.  
 DR PROSITE: PS50076; DnaJ\_2; 1.  
 SO SEQUENCE 796 AA; 87577 MW; B2E107AB32DF9986 CRC64;

Query Match 7.2%; Score 124; DB 10; Length 796;  
 Best Local Similarity 23.6%; Pred. No. 0.017;  
 Matches 68; Conservative 36; Mismatches 132; Indels 52; Gaps 11;

QY 74 TPNAQ-----PFLEKQPVHSGSKILPRESKGVHVELRSPTHGSOOSRK 120  
 DB 164 TPQSQSQSPQHNOQSQSPQHNPNOQTVSFQNOQLP-FOLHQPFAAVQTSQSQSQRS 222  
 QY 121 VPDYHVTMOOLESTRNGEPHILACD-----RPSKHLKLVFIRHCLRALRLPRISIDM 175  
 DB 223 QFDQQLLDQDQASMRQOVGOEHSQSSGDMKRPVEEVRLINLNATERVQSSRRFAEPE 282  
 QY 176 ESPPLNSGEALSPATAKDEIT-----OMILKSAARSEL--GMVYSKROEYFLRRARR 228  
 DB 283 RPPRSRSPFG-----SHRTPSTELTWASKPTPVSEPVHSELVPMQISEPAQYOL--SSRS 337  
 QY 229 RRFAMKPVLOSISEMVMVEFTPMAYRDSG-----SPKKNASTSLPGKNISPPROVSV 284  
 DB 338 SEAAQLSLPVSQSS-----HASQTRSNQSHAVSKPQVSKRPHPFPSPQPP--TSN 390  
 QY 285 PQRSSPPKNSP-----PPQAFVARTASKYSAAQQOVQNR 322  
 DB 391 PPPLSQPPSNKPPPMQSSQSNKPPPVQSSQSKPLVQSSQSRK 438

RESULT 5  
 Q9XEO2 PRELIMINARY: PRT: 1104 AA.  
 AC 09XEO2:  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL.125.9 KDA PROTEIN.  
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC panicoidae; Andropogoneae; Sorghum.  
 OC NCBI\_TaxId=4558;  
 QX  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Liaca V., Lou A., Young S., Messing J.;  
 RT "Retriectransposable elements of Sorghum bicolor.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF114171; AAD27567.1;  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 KW Hypothetical protein.  
 SO SEQUENCE 1104 AA; 125894 MW; 59A1FE2CB899C855 CRC64;

Query Match 7.1%; Score 122.5; DB 10; Length 1104;  
 Best Local Similarity 33.0%; Pred. No. 0.035;  
 Matches 37; Conservative 17; Mismatches 41; Indels 17; Gaps 5;

QY 228 RRFAMKPVLOSISEMVMVEFTPMAYRDSGSPKKNAST-PSLPGKNISPPROVSV 284  
 DB 548 RRRFLTPQHDDSSNLSPLISPAHNSAAGGNAGSPPTPAARPTLPFP-----PPRSPP 603  
 QY 285 PQRSSPPKNSP-----PPQAFVARTASKYSAAQQOVQNRGNK 326  
 DB 604 PPRSPSPPSPTPPRPSPTSKRSKAPPPPPAPPSPKASVSPSPKRSK 655

RESULT 6  
 Q70495 PRELIMINARY: PRT: 897 AA.  
 ID 070495  
 AC 070495;





Query Match 6.4%; Score 111; DB 10; Length 202;  
Best Local Similarity 21.9%; Pred. No. 0.041;  
Matches 54; Conservative 24; Mismatches 89; Indels 80; Gaps 7;

```
QY 16 RYVYPDGHEVRLKSCSYHDLGNDYVCGSTP-----YITTRMA--AEVL 63
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 19 RYVLSGNGVONLEETVAEMLENOHYVEFPDSSISFNNDKATVKKRLAPLADPTL 78
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 64 EGVATYFVCATPNAPFLEROPKVVHRSKILPRFSKHGVHRELSPTHSGSQSRKVF 123
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 79 EBGKRYLV-----LPA-KRSGGKAAKSSSAVLTESEMKKML 113
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 124 YHSVYMOQLESIRNEGPEPHLACDRSKHLKLVFRICLRALRLPRISIDLMESPLNLS 183
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 114 -----FSATYAVHSSFSYEGILPFT 135
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 184 GEAISPATAKDEITOMILKSAARSELGMVYSKROEFLRRARRRKFAMKPVLOSISEM 243
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 136 TRSYKNNRPATD--TVAAATSVGRLAEKMEEDREPLSRQLSGR--GMRPSLDPIREK 190
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 244 KPVMEFH 250
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 191 KAKKKH 197
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

```
RESULT 11
Q91T14 PRELIMINARY; PRT; 393 AA.
AC 091T14:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE T117.
OS Tupia herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae.
OX NCBI_TaxID=10397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2;
RC MEDLINE=21211637; PubMed=11312357;
RA Bahr U., Darel G.;
RT "Analysis and Characterization of the Complete Genome of Tupia (Tree
RT Shrew) Herpesvirus."
RL J. Virol. 75:4854-4870(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2;
RA Darel G., Bahr U.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF281817; AAK57163.1; -.
SQ SEQUENCE 393 AA; 41784 MW; F48A27B83C53BDF CRC64;
```

Query Match 6.3%; Score 109.5; DB 12; Length 393;  
Best Local Similarity 22.7%; Pred. No. 0.13; Indels 75; Gaps 9;  
Matches 45; Conservative 25; Mismatches 53;

```
QY 140 PEPHLAGDRPRSKH-----LKLVEIR-----HCL-RALRLPRISIDLMESPLP 180
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 68 PPPPEFGDPAHNGTTPPTVTITVTYVPPASPPIHAYPRHNRSP-----ESPSP 120
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 181 NLSGE-----ALSPATAKDEITOMILKSAARSELGMVYSKROEFLRRARRR 229
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 121 SEPTAEITAAIAVSAPRALGPAITEYEGSLHPRKRSPTTAETGALLE----- 167
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 230 KFAAMKPVLOSISEMKPVMEFHTPMAYRDSGSPKNASTPSLPGKNISPPROQVSPQRS 289
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 168 --VLKPIEAVSGT-----QTAPDPTPTPR--SPR--SPPRAI 202
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 290 PPPKNVSPPOPAFYART 307
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

DB 203 PPPPPPPPPPPPAET 220

```
RESULT 12
Q9FHX1 PRELIMINARY; PRT; 1073 AA.
AC 09FHX1:
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE TWY RESISTANCE PROTEIN-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RC MEDLINE=99397451; PubMed=10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT pl and TAC clones."
RL DNA Res. 6:183-195(1999).
DR EMBL; AB017067; BAB08447.1; -.
DR InterPro: IPR000345; CYCC_heme_bind.
DR InterPro: IPR000408; RCC1.
DR InterPro: IPR000306; znf_FYVE.
DR Pfam: PF01363; FYVE; 1.
DR Pfam: PF00415; RCC1; 6.
DR PRINTS; PRO0633; RCCNDSATTON.
DR SMART; SM00064; FYVE; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS50012; RCC1_3; 6.
SQ SEQUENCE 1073 AA; 116497 MW; 7A9F6C2F5DA40CD4 CRC64;
```

Query Match 6.3%; Score 109.5; DB 10; Length 1073;  
Best Local Similarity 23.6%; Pred. No. 0.49;  
Matches 89; Conservative 46; Mismatches 147; Indels 95; Gaps 16;

```
QY 1 MGLTPESC-----TVQGVRRVYPPDGHV-----ENLSKSCSYHDLGNDYVCGSTP 50
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 507 VGLTTSCKXYTMGSTVYGGGIGNPNADGKLPCLVEDKLTIDC--VEEIAAGAYHVAVLTRN 565
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 51 YITNRMAAEVLEYGVTYFVCATPNAPFLEROPKVVHRSKILPRFSKHGVHRELRS 110
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 566 EVFTWKGANGRGRGHGVDEDRKAPTLVDALKERHYKNINAGSGNFAALCLH-----KWS 620
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 111 PTHSGQGS--RKVEDY-----HSVTMOQLESTR-----NKGPEPHLAGDRPS 150
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 621 GTEQSQCSACROAFGFTPRKRNHCYNGGLVHSCSSKSLKALAAPNNG--RPRVCD--S 677
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 151 KHLVFIIRICLALRLPRISIDLMESPLNLSGE-----ALSPATAKDEIT 198
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 678 CHSL-----SKVSEANIDSKKNVPRLSGENKDRLDTEIRLAKSGIPSIDLIK 728
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 199 OMILKSAARSELGMVYSKROEFLRRARRRKFAMKPVLOSISEMKPVMEFHTPMAYRDS 258
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 729 QLDNRRAAROK-----KADTESLVRTSQPLQLKDALTNVADL----- 768
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 269 GSPPKNSTSTSLPGKNISPPROVSVPPRSSPPKKNVSPPO-----PAFYARTASKYS 312
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 769 RGPKPAPVTP-----SSSRPVSFRRSSPPRSVTPPLPVNGLGFTSIABSLKTN 820
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 313 -AASQOYQRRNGNAKSL 328
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 821 ELNQEVVRLRAQESL 837
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
RESULT 13
```

Q9UHA8  
ID Q9UHA8 PRELIMINARY; PRT; 2296 AA.  
AC Q9UHA8  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE SPLICING COACTIVATOR SUBUNIT SRM300.  
GN SRM300.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20132238; PubMed=1066804;  
RA Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,  
RA Rosonina E., Sharp P.A.;  
RT "The SRM160/300 splicing coactivator subunits.";  
RL RNA 6:111-120(2000).  
DR EMBL; AF201422; AAF21439.1;  
DR InterPro: IPR002965; P. rich-extendens.  
DR PRINTS: PRO1217; PRICHEXTENS.  
SQ SEQUENCE 2296 AA; 251964 MW; 17C0BD4EA10A9CF9 CRC64;

Query Match 6.3%; Score 109.5; DB 4; Length 2296;  
Best Local Similarity 26.5%; Pred. No. 1.3;  
Matches 53; Conservative 27; Mismatches 75; Indels 45; Gaps 10;

OY 162 LRALRLPRISIDIMESPLNGEALSPATAKDEI---TQML-KSANSSEIGMTYVK 216  
DB 279 LEVQLQKLIQPLQGEVLPLOGDGREMLSVNQVLPSTORASSPEATQPSPEDK 338  
OY 217 ROEFYLRARRRRKFAKMP-VLOSISEMKPYMEFHTP-MAYRDSGSPKNASTSLGPK 274  
DB 339 DKD-----KKKSATRTSPSPERSGTGPPAPTPPLAERHSGSPQPLATTLTSGEP 390  
OY 275 NISPPROSVPORSSPPPKN-----VSP-POPAFPAKAS----- 309  
DB 391 -VNPFSAS-PTRDSPPKSPKPEKLPOSSSSBSPSPQPKVSRHASSSPSPKAPAPAG 448  
OY 310 ---KYSASQOVORNRGNK 326  
DB 449 SHREISSSPKSKRSHGRK 468  
RESULT 14  
O9M2X4  
ID O9M2X4 PRELIMINARY; PRT; 651 AA.  
AC O9M2X4  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE HYPOTHETICAL 72.3 KDA PROTEIN.  
GN T16K5.190.  
OS Arabidopsis thaliana (mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX EU Arabidopsis sequencing project;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL132965; CAB66922.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 651 AA; 72344 MW; F4ED1138046F2FID CRC64;

Query Match 6.3%; Score 109; DB 10; Length 651;  
Best Local Similarity 22.5%; Pred. No. 0.28;  
Matches 62; Conservative 30; Mismatches 92; Indels 92; Gaps 13;

OY 52 TITNRMAAEVLEVGVTFVCATPNAQPLERQPKVHNGS--KILRFKHCYHREL 109  
DB 387 SISRKILAQRIEODDEFECS-----PGLIHTSFASPLPGSKKLRVPLS 435  
OY 110 -----SPHSGQSKRYVDHYVTMOQLESIRNEGPEPHLAGDRPSKHLKLVIRHCLR 163  
DB 436 ESDIPSSFHKTPPEGRR-----KLTKELQKFTKESTE-----N 469  
OY 164 ALRLPRISIDIMESPLPLNGEALSPATAKDEITQMLKSAAR---SEIGMYVSKROF 220  
DB 470 ALK-----ELVSSPEBSSSGSE-----KQAAAGTSEPVGECTKKKKQ- 506  
OY 221 YLRARRRRKFAKMPVLOSISEMKPYMEFHTP---AYRDSGSPKNASTPS-LDGPKN 275  
DB 507 -----RVRYKERNKRFSDKSTYRKKKMSYODQHVPYSAPPPQGYPPREGYPGAPPPAG 561  
OY 276 ISPPR--QVSVFORSSPPPK-----NVSPPO 300  
DB 562 YPPPYPPQAGYPPAGYPPPOGGYGGYPPAGYPPQ 597  
RESULT 15  
O9EEF8  
ID O9EEF8 PRELIMINARY; PRT; 517 AA.  
AC O9EEF8  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE HYPOTHETICAL P60 PROTEIN  
OS Anticarsia gemmatilis nuclear polyhedrosis virus (AgMNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=31507;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92268862; PubMed=1588315;  
RA Zanotto P.M., Sampaio M.J., Johnson D.W., Rocha T.L., Marunjak J.E.;  
RT "The Anticarsia gemmatilis nuclear polyhedrosis virus polyhedrin gene  
region: sequence analysis, gene product and structural comparisons.";  
RL J. Gen. Virol. 73:1049-1056(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN-2D;  
RA Zanotto P.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN-2D;  
RX MEDLINE=94044871; PubMed=8228320;  
RA Zanotto P.M.A., Kessing B.D., Marunjak J.E.;  
RT "Phylogenetic interrelationships among baculoviruses: Evolutionary  
rates and host associations.";  
RL J. Invertebr. Pathol. 62:147-164(1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX STRAIN-2D;  
RX MEDLINE=20033815; PubMed=10564750;  
RA Chen X., Ijkel W.F.J., Dornay C., Zanotto P.M.A., Hashimoto Y.,  
RA Faktor O., Hayakawa T., Wang C.H., Premunekar A., Krell P.J., Hu Z.,  
RA Vlak J.M.;  
RT "Identification, sequence analysis and phylogeny of the Ief-2 gene of  
RT Helicoverpa armigera single-nucleocapsid baculovirus.";  
RL Virus Res. 65:21-32(1999).  
DR EMBL; Y11753; CAC03567.1; -  
DR InterPro: IPR003124; WH2.  
DR Pfam: PF02205; WH2; 2.  
DR SMART: SM00246; WH2; 2.







## Blast 2 Sequences results

PubMed

Entrez

BLAST

OMIM

Taxonomy

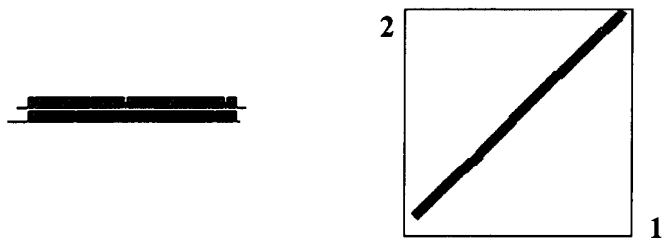
Structure

### BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.3 [Apr-24-2002]

Matrix BLOSUM62  gap open: 11 gap extension: 1  
 x\_dropoff: 50 expect: 300. wordsize: 3

**Sequence 1** gi 2851522 Mitogen-activated protein kinase 12 (Extracellular signal-regulated kinase 6) (ERK-6) (ERK5) (Stress-activated protein kinase-3) (Mitogen-activated protein kinase p38 gamma) (MAP kinase p38 gamma). **Length** 367 (1 .. 367)

**Sequence 2** gi 1136298 WIPK [Nicotiana tabacum] **Length** 375 (1 .. 375)



NOTE: The statistics (bitscore and expect value) is calculated based on the size of nr database

Score = 292 bits (747), Expect = 5e-78  
 Identities = 149/339 (43%), Positives = 219/339 (63%), Gaps = 8/339 (2%)

```

Query: 21  WEVRAVYRD-LQPVGSGAYGAVCSAVDGRGAKVAIKKLYRPFQSELF AKRAYRELRLLK 79
      +E+  YR  + P+G GAYG VCS ++      VA+KK+  F  + AKR  RE++LL+
Sbjct: 36  FEITTKYRPPIMPPIGRGAYGIVCSVLNTELNMVAVKKIANAFDIYMDAKRTLREIKLLR 95

Query: 80  HMRHENVIGLLDVFTPDETLDDFTDFYLVMPFMGTDLGKLMK-HEKLGEDRIQFLVYQML 138
      H+  HENVIGL DV  P      +F+D Y+      M TDL ++++ ++ L ED  Q+ +YQ+L
Sbjct: 96  HLDHENVIGLRDVIPPP-LRREFSDVYIATELMDTDLHQIIRSNQGLSEDHCQYFMYQLL 154

Query: 139 KGLRYIHAAGIIHRDLKPGNLAVNEDCELKILDFGLARQ--ADSEMTGYVVTRWYRAPEV 196
      +GL+YIH+A ++HRDLKP NL VN +C+LKI DFGLAR      + MT YVVTRWYRAPE+
Sbjct: 155 RGLKYIHSANVLHRDLKPSNLLVNANCDLKICDFGLARPNIENENMTEYVVTRWYRAPEL 214

Query: 197 ILNWMRYTQTVDIWSVGCIMAEMITGKTLFKGSDHLDQLKEIMKVTGTPPAEFVQRLQSD 256
      +LN  YT  +D+WSVGC I  E++ K LF G DH+ Q++ + ++ GTP      + LQ++
Sbjct: 215 LLNSSDYTAAIDVWSVGCIFMELMNRKPLFGGKDHSVHQIRLLTELLGTPTEADLGFLQNE 274

Query: 257 EAKNYMKGLEPELEKKDFASILTNASPLAVNLEKMLVLDAEQRV TAGEALAHYPFESLHD 316
      +AK Y++ LP+ ++  A +  + +PLA++L++KML D  +R+T  EAL HPY  LHD
Sbjct: 275 DAKRYIRQLPQHPRQQLAEVFPHVNPLAIDLVDKMLTFDPTTRITVEEALDHPYLAKLHD 334

Query: 317 TEDEPQVQKYDDSFDDVDRTL--DEWKRVTYKEVLSFKP 353
      DEP +      SFD  + + ++ K + Y+E LS  P
Sbjct: 335 AGDEP-ICPVPF SDFEQQGIGEEQIKDMIYQEALSLNP 372
  
```

CPU time: 0.08 user secs. 0.01 sys. secs 0.09 total secs.



Lambda	K	H
0.320	0.137	0.405

## Gapped

Lambda	K	H
0.267	0.0410	0.140

Matrix: BLOSUM62

Gap Penalties: Existence: 11, Extension: 1

Number of Hits to DB: 1022

Number of Sequences: 0

Number of extensions: 86

Number of successful extensions: 4

Number of sequences better than 300.0: 1

Number of HSP's better than 300.0 without gapping: 1

Number of HSP's successfully gapped in prelim test: 0

Number of HSP's that attempted gapping in prelim test: 0

Number of HSP's gapped (non-prelim): 1

length of query: 367

length of database: 330,426,180

effective HSP length: 123

effective length of query: 244

effective length of database: 219,683,991

effective search space: 53602893804

effective search space used: 53602893804

T: 9

A: 40

X1: 16 ( 7.4 bits)

X2: 129 (49.7 bits)

X3: 129 (49.7 bits)

S1: 41 (21.8 bits)

S2: 60 (27.7 bits)

